Vol. 39, No. 5 October, 1949

Journal of Experimental Psychology

EDITED BY

FRANCIS W. IRWIN

UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CLARENCE H. GRAHAM, *Columbia University*
HARRY HELSON, *Brooklyn College*
DONALD B. LINDSLEY, *Northwestern University*
DONALD G. MARQUIS, *University of Michigan*
GEORGE A. MILLER, *Harvard University*
L. L. THURSTONE, *University of Chicago*
G. R. WENDT, *University of Rochester*
JOSEPH ZUBIN, *N. Y. Psychiatric Institute, Columbia University*

CONTENTS

An Analysis of the Superiority of Binocular over Monocular Visual Acuity: M. W. HOROWITZ... 581
The Electrical Response of the Human Retina During Dark-Adaptation: E. P. JOHNSON..... 597
Effects of Brief Flashes of Light upon the Course of Dark Adaptation: D. A. GRANT AND F. A. MOTE 610
Experimental Studies in Thinking: I. Scattered Speech in the Responses of Normal Subjects to Incomplete Sentences: N. CAMERON AND ANN MAGARET..... 617
Various Degrees within a Single Drive as Cues for Spatial Response Learning in the White Rat: R. BLOOMBERG AND W. B. WEBB 628
Reinforcement and Extinction as Factors in Size Estimation: W. W. LAMBERT, R. L. SOLOMON, AND P. D. WATSON 637
Semantic and Phonetographic Generalizations of Salivary Conditioning to Verbal Stimuli: G. RAZRAN 642
Studies in Spatial Learning: VII. Place and Response Learning under Different Degrees of Motivation: E. C. TOLMAN AND H. GLEITMAN..... 653
Sign Learning as a Factor in Extinction: J. P. SEWARD AND N. LEVY..... 660
Retention and Warming-Up Effects in Paired-Associate Learning: A. L. IRION..... 669
Repetition of Two Basic Experiments on Reminiscence in Serial Verbal Learning: C. E. BUXTON 676
Attempted Pupillary Conditioning at Four Stimulus Intervals: E. R. HILGARD, C. E. DUTTON, AND J. S. HELMICK 683
'Spread' in Recalling Failure and Success: J. NUTTIN..... 690
The Perception of the Vertical: II. Adaptation Effects in Four Planes: G. E. PASSEY AND F. E. GUEDRY, JR. 700
Permanence of Experimentally Induced Changes in the Attractiveness of Activities: MILDRED E. GEBHARD 708
Experimental Studies in Rote-Learning Theory: VIII. Distributed Practice of Paired Associates with Varying Rates of Presentation: C. I. HOVLAND..... 714
The Construction of Subjective Brightness Scales from Fractionation Data: A Validation: R. M. HANES 719
The Effect of Psychophysical Method and Context on Pitch and Loudness Functions: J. M. DOUGHTY 729

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. & R., authorized September 11, 1947.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published bi-monthly by the American Psychological Association, Inc. The annual subscription rate is $7.00 (foreign, $7.50); single copies, $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Articles of a length exceeding 25 printed pages cannot be accepted.

With rare exceptions, articles are published in the order of their receipt. Authors are supplied with 50 free offprints, without covers, in the case of articles of seven or more printed pages, and with five free copies of the issue in the case of shorter contributions. One-half of the cost of tables, cuts, formulas, and author's alterations in galley proof is charged to the author.

Immediate publication can be given to articles whose authors assume the full cost of publication.

Address articles submitted for publication to the editor, Francis W. Irwin, Department of Psychology, University of Pennsylvania, Philadelphia 4, Pa.

This JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. ANDERSON and W. L. VALENTINE, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, **41**, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *E*s, *E*'s, and *E*s'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts should be typed double-spaced throughout, *including tables, footnotes, and bibliography*. Tables, figures, and footnotes should be placed on separate pages at the end of the manuscript.

Journal of
Experimental Psychology

VOL. 39, NO. 5 OCTOBER, 1949

AN ANALYSIS OF THE SUPERIORITY OF BINOCULAR OVER MONOCULAR VISUAL ACUITY *

BY MILTON W. HOROWITZ

University of Kansas

In nearly all cases, the measurement of binocular visual acuity yields a lower threshold than monocular visual acuity. This fact has long been known, and our own data,[1] on several occasions, have confirmed it. Binocular acuity is found to be significantly better than 'best eye' where 'best eye' is taken in each judgment as the better monocular judgment of either eye. The problem is simple in its fact but not in its theory; it is not clear why the superiority should exist.

An examination of the usual clinical situation affords a basis for evaluating binocular and monocular vision. Monocular vision differs from binocular in that different fields are presented to the two eyes and judgments are required concerning features present in one of the fields while the other is ignored. Conventionally, an acuity pattern is used for the tested eye; the non-tested eye is occluded by a blind which presents both a low intensity and an unfigured field. Obviously, such treatment does not prevent the participation of the non-tested eye in the visual act, although this fact is often overlooked. It is in this situation that the superiority of binocular over monocular visual acuity must be interpreted.

This problem is important both in its clinical, practical aspects and in its contribution to visual theory. Although there is still controversy as to whether summation occurs for the light threshold, the facts are clear that binocular acuity does exceed monocular acuity. The implications for a theory of interaction are apparent.

Nevertheless, many hypotheses have been advanced which disregard an explanation in terms of interaction and which attempt to explain binocular superiority in terms of external, or peri-

* The experimental work reported here formed part of a contract with the Air Materiel Command, Wright Field, Dayton, Ohio, and was done in the Vision Laboratory, Department of Physiology, School of Medicine, Stanford University.

The paper is an abridgment of a dissertation submitted in partial fulfillment of the degree of Doctor of Philosophy, in the Department of Psychology of Stanford University. The writer wishes to express his sincere appreciation to Professor Frank W. Weymouth who guided the study, and to Professors Donald W. Taylor, Quinn MacNemar, and Monroe Hirsch for their many helpful suggestions and stimulating criticisms throughout the course of the experiment.

[1] Vision Laboratory, Stanford University. Unpublished data.

pheral, or even fortuitous factors. It has been said, for example, that binocular judgments are more easily made; there is greater freedom from fatigue in binocular vision. It has also been said that the increase in retinal area is responsible. A rather obvious reason advanced for binocular superiority (9) is that optical defects frequently do not appear in the two eyes in the same manner or to the same degree. Astigmatism often fulfills these conditions in that it may be less severe in one eye or may occur in a different axis than in the other eye. Consequently, binocular vision always capitalizes on this, and for this reason alone, some superiority should exist.

It has been claimed that fixation is better in binocular vision and that this plays a role in the superiority of binocular acuity. Better binocular accommodation has also been offered as an hypothesis, either by itself, or associated with convergence. Quite recently, both Pirenne (42) and Bárány (9), the former working with the light threshold and the latter with visual acuity, have advanced a statistical hypothesis. If we assume both eyes act independently, then by chance, at the threshold, when using two eyes, one will generally be 'good' when the other is 'bad' and binocular acuity must necessarily be better than monocular. Essential to the argument, of course, is a range of acuities for the two eyes.

Not all of these hypotheses merit critical attention. Some are but speculations unsupported by data. Of the more plausible hypotheses, it would seem that optical defects, such as astigmatism, must contribute to binocular superiority. It is, however, impossible to estimate its significance or the amount of the contribution, but it is planned to evaluate this variable thoroughly in a future paper.

I know of no evidence showing binocular fixation to be better than monocular, but, even if it were, it is difficult to understand wherein this would aid acuity. Logically, eye movements should serve to give a graded perception and better definition, whereas better fixation should serve to increase adaptation and the incidence of after-images both of which would lower acuity (6, 7, 39).

Binocular accommodation is known to exceed monocular and, at the reading distance, it is quite possible that accommodation and convergence would aid binocular acuity. At six meters,[2] the usual visual acuity testing distance, both accommodation and convergence are reduced to a negligible amount.

Our own data have failed to support the statistical hypothesis of Pirenne and Bárány when applied to visual acuity data. When the two eyes of an individual are markedly different, the distribution of acuities for the two eyes should scarcely overlap, and binocular acuity should be no better than 'best eye.' Unpublished data secured in the Vision Laboratory of Stanford University, however, have shown that it is better. Two individuals, both almost amblyopic in one eye, have given binocular acuities higher than the better monocular acuity.

The present hypotheses.—The hypotheses for the experiment reported were drived from an analysis of the differences between binocular and monocular vision. There are three important factors here:

1. In monocular vision, both eyes participate in the visual act, although one eye is occluded.

2. The occluded eye is presented with a low intensity relative to the non-occluded eye.

3. The occluded eye is presented with an unfigured field whereas the non-occluded eye views the acuity pattern.

These factors form the basis for two hypotheses which are tested in this experiment:

[2] The data showing binocular superiority are at approximately the six-meter distance. There is only $\frac{1}{6}$ of a diopter of accommodation left at this distance.

1. The low intensity of the occluded eye will increase the diameter of the pupil of the non-occluded eye resulting in a lowered monocular acuity.

2. The difference in intensity and pattern in the two eyes will result in either rivalry or fusion of the dissimilar fields, leading to a reduction in monocular acuity.

Pupil size is important in visual acuity, for it controls partially the intensity of light impinging upon the retina. It may also reduce the effect of refractive errors. That is, the smaller pupil present in binocular vision decreases the pencil of rays and reduces the size of the diffusion circle. Even though the brightness of the retinal image is decreased, a clearer image may be obtained and visual acuity will be increased.

We may also expect that brilliance and pattern conflict will ensue in monocular vision. The field of the occluded eye differs from the field of the non-occluded eye in both brilliance and pattern. Rivalry, in the narrow sense of a temporal fluctuation of the two fields, may occur; or, through a central fusion of the two fields, there may be a reduction in contrast. It is apparent that either the pattern or the brilliance conflict could reduce monocular acuity. And, since visual acuity is essentially a brightness discrimination (**1, 11, 24**), it is equally apparent that a reduction in contrast between figure and ground would also lower monocular acuity.

Apparatus and Method

Four pieces of equipment are described here. These include two test objects, an apparatus to control the intensity of light in the occluded eye during monocular judgments, and the artificial pupil. The two test objects used were the Ives grating and the Landolt broken ring. The Ives visual acuity test object has been well described in the literature (**34, 35**). It consists essentially of two glass plates ruled with fine, parallel, equidistant lines to form a grating. If two such plates are overlaid and rotated, parallel dark bands appear which grow broader and more distantly separated as the angle between the gratings increases. As Ives has remarked, the test object is such that it is continuously variable in size and has the attractive feature of maintaining constant the total flux entering the eye from the object.

The Landolt broken ring was projected on a ground glass screen by a Clason Visual Acuity Meter (Bausch and Lomb). The test object was photographed from printed models made to specification and slides were prepared. The Clason, like the Ives, allowed the broken ring to be continuously variable in size. The rack and pinion arrangement of the Hamilton slide furnished with the apparatus allowed the gap of the broken ring to be presented in any position desired. Hence, the two test objects were comparable in all adjustable respects. Both objects gave decimal acuities equivalent to the comparable Snellen fraction. Both were admirably suited for obtaining the threshold of resolution.

An apparatus was constructed[3] to present varying intensities of light to either eye while the other eye was viewing the test object. Essentially it consisted of a light box enclosing four six-watt white fluorescent bulbs directed into a mirror at 45°. The light was diffused through opal flash glass before entering the eye. The flash glass, placed at approximately 2.5 cm. from the *S*'s eye, served the further purpose of presenting no point of fixation, thereby decreasing the possibility of accommodation. Within the light box were two slots for ground glass filters to diffuse the light within the box. Over the ground glass plates could be placed a prepared, graded series of paper filters of any density desired. The whole was mounted upon a sturdy frame allowing the *S* to place himself immediately before the light source with either eye. With a shift to either side of a few inches binocular judgments could be made. An adjustable chin rest to immobilize the head was attached to the framework and a removable eyepiece was attached to the light box.

The only other apparatus was the artificial pupil which allowed for apertures varying in one-mm. steps from one mm. to seven mm. The testing laboratory was a room 19.5 meters long, separated into two sections by a partition. The Ives grating, the Clason Acuity Meter, and the ground glass screen were placed in one section. The room in which the *S* was seated was 8.7 meters long, 2.0 meters wide, and 2.6 meters

[3] This apparatus was constructed by Mr. Milton Waldman of the Department of Physiology, Stanford University.

high. This room was bare with the exception of two built-in tables along the left wall. As viewed by the *S*, the partition and adjacent walls presented a uniform white surface illuminated to approximately 38.5 ml.[4] A seven-cm. circular aperture in the partition 1.3 meters from the floor permitted a view of the test object.

Design of the experiment.—It has been hypothesized that brilliance differences and pupil changes account for much of the superiority of binocular over monocular visual acuity. These two hypotheses require two main variables, *intensity* and *pupil changes*, in the design of the experiment.

Four intensities were selected for presentation to the *S*'s occluded eye during monocular judgments and represent the experimental control of brilliance differences. These four intensities were 195 ml., 40 ml., 0.29 ml., and 0.02 ml. The highest intensity possible in the light box was 195 ml., the first condition. The fourth condition, 0.02 ml., was a low intensity and one which would yield monocular judgments closely comparable to the clinical situation. The 40-ml. condition represented approximately the same brilliance as the walls and the test objects, which, on logical grounds, should reduce brilliance differences to a minimum. The 0.29-ml. condition was used to obtain an acuity judgment intermediate between that for 40 ml. and 0.02 ml. Pupil changes, of course, were controlled by the artificial pupil.

Positions, *eyes*, and *subjects* are also variables in this design. Four positions of the test object were used. Varying the position of the test object made the acuity judgments more objective and, when used with equal frequency, prevented the average acuity from being influenced by astigmatism. For comparison with binocular judgments, it was necessary to test the eyes separately. Consequently, the eyes of the *S* represented another variable, although one not immediately concerned in the primary analysis. *Subjects*, also, is a variable since 10 individuals comprised the sample tested. Individual differences in acuity are to be expected, however, and this variable, like *eyes*, is not of particular interest in the design of the experiment. It was finally decided, in order to enhance the inductive value of the experiment, to use two different test objects, the grating and the broken ring. *Test object*, therefore, enters as another variable in the design of the experiment.

In view of the large number of variables, an analysis of variance was considered the appropriate statistical technique. However, since six variables represents a cumbersome analysis, and since many of the resulting interactions would be of doubtful importance and interpretation, it was decided to design four experiments which would include all variables. These four different, though related experiments are: an experiment with pupil changes controlled for each test object and an experiment with pupil changes uncontrolled for each test object. The main variables used in each of the four experiments are ten *subjects*, four *intensities*, four *positions*, and two *eyes*, as may be seen in the design of the analysis in Table I.

TABLE I

DESIGN OF THE ANALYSIS OF VARIANCE

Sum of Squares	Degrees of Freedom
Subjects	9
Intensity	3
Positions	3
Eyes	1
S × I	27
S × P	27
S × E	9
I × P	9
I × E	3
P × E	3
S × I × P	81
S × P × E	27
I × P × E	9
S × I × E	27
S × I × P × E	81
Error	1280
Total	1599

The three experimental variables were randomized. Binocular judgments were not randomized, since they did not enter into the same analysis as the monocular judgments. It was felt necessary only to insure that the binocular judgments would be affected by neither practice nor fatigue effects. Accordingly, a series of binocular judgments were inserted at pre-arranged intervals throughout the randomized monocular series. This has, of course, led to some control that could have been eliminated by randomizing all series, both monocular and binocular.

Experimental procedure.—The *S* was seated comfortably at six meters from the test object, the chair raised or lowered to accommodate for differences in height. The chin rest was adjusted to maintain the head steady, and a few words of instruction given. The *S* was told to immobilize his head as much as possible (some free movement was possible with the head in the chin rest) and to call as soon as the test object could be resolved. Numbers of one through four were assigned to the four positions of the

[4] All light readings were taken with the Macbeth Illuminometer. In each case, 10 readings were taken and averaged.

test object. If the *S* called the wrong position, *E* continued to enlarge the pattern at a constant rate (approximately 10–12 sec. per judgment) until it was correctly resolved.

Each presentation of the test object was announced by a buzzer signal which served to maintain alertness. Short rests of 1.5 to 2 min. were given after each series of 20 judgments under one condition and longer rests of 10–20 min. were given if the *S* complained of fatigue. After each series of 20 judgments a different intensity, a different eye, or both, were used for the next series of 20 judgments. In all, 160 monocular judgments (20/intensity/eye) and 40 binocular judgments constituted the testing for one test object. This was repeated for the same test object with pupil changes controlled by an artificial pupil (a two-mm. aperture) and the same repeated for the other test object.

Results

The analysis of variance.—The primary treatment of the results is the analysis of variance. In Table II may be seen the analysis for one of the four experiments (the Ives grating with the normal pupil). The explanation of the results shown in this table will indicate the primary statistical treatment for all four experiments. Reference to Table I may be helpful in this discussion. Table II shows the source of variance, sums of squares, degrees of freedom, variances, *F* ratios, and probability level for all variables and interaction terms listed. The primary variables are *subjects* with nine *df*, *intensity* with three *df*, *positions* with three *df* and *eyes* with one *df*. All other interactions are combinations of the primary variables with the appropriate *df*.

The error term represents variability of *S* and is the error of measurement. For each *S*, within each eye and each intensity and for each position, five separate judgments were made. The variance due to these five judgments is error.

Since the error term does represent an error of measurement, it is the

TABLE II

Analysis of Variance for the Grating with the Normal Pupil

Source	Sum of Squares	*df*	Variance	*F*	5%	1%
Subjects	70.0590401	9	7.78434	1898.620	1.88	2.41
Intensities	1.7422872	3	0.58076	141.649	2.60	3.78
Positions	14.0559767	3	4.68533	1142.763	2.60	3.78
Eyes	0.9985006	1	0.99850	243.537	3.84	6.64
S × I	1.0547047	27	0.03906	9.527	1.52	1.79
S × P	13.2576702	27	0.49102	119.761	1.52	1.79
S × E	1.2951175	9	1.14390	35.098	1.88	2.41
I × P	0.0843346	9	0.00937	2.285	1.88	2.41
I × E	0.1371362	3	0.04571	11.149	2.60	3.78
P × E	0.7603597	3	0.25345	61.817	2.60	3.78
S × I × P	0.6359160	81	0.00785	1.915	1.28	1.41
S × P × E	3.3004997	27	0.12224	29.815	1.52	1.79
S × I × E	0.6187782	27	0.02292	5.590	1.52	1.79
I × P × E	0.0883185	9	0.00981	2.393	1.88	2.41
S × I × P × E	0.6257796	81	0.00773	1.885	1.28	1.41
Error	5.2460800	1280	0.00410			
Total	113.9604995	1599				
Intensities	1.7422872	3	0.58076	14.868	2.96	4.60
Positions	14.0559767	3	4.68533	9.542	2.96	4.60
Eyes	0.9985006	1	0.99850	6.939	5.12	10.60
S × I	1.0547047	27	0.03906			
S × P	13.2576702	27	0.49102			
S × E	1.2951175	9	0.14390			

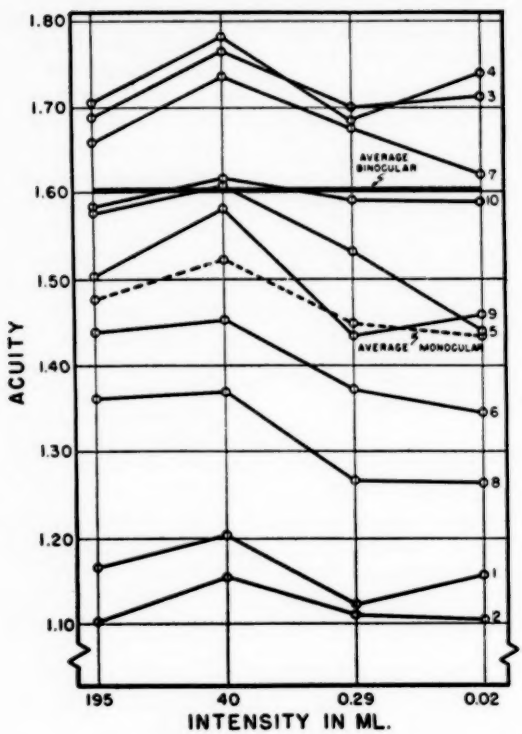


FIG. 1. The relationship of acuity to the four conditions of intensity for the Ives grating and the normal pupil

Figs. 1–4 show the acuity levels achieved by the 10 *S*s under the four experimental conditions of intensity. The average binocular acuity for all *S*s under standard conditions of approximately 40 ml. is given in each figure by the heavy line; the average monocular acuity by the dotted line. The numbers to the right of each figure indicate the *S*. Since the same *S*s were used in all four experiments, the acuity of each can be followed through the conditions of different test object and the normal and artificial pupil.

proper term with which to evaluate *subjects*. The *F* ratio for *subjects* in all four experiments is highly significant. This is not surprising since the acuities of the different *S*s differed to a marked extent. The other three primary variables are evaluated by the first order interactions relating to them. Since generalizations are to be made to the population of *S*s, it is proper to use the first order interactions concerned with *S*s to test the primary variables. The appropriate error terms are S × I, S × P, S × E.[5]

When evaluated in this manner, *intensity* is also significant beyond the

[5] Table II is in two sections. The upper section represents the *F* ratios of all variables and interactions with error. The lower section represents the *F* ratios of *intensities*, *positions*, and *eyes* with the first order interaction applicable in each case. The *F* ratios of the primary variables in the upper section represent results that are valid for the subjects used in this study.

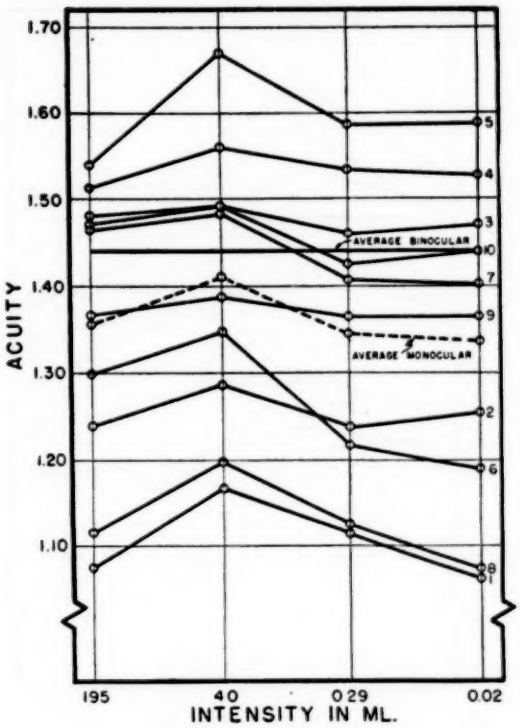


FIG. 2. The relationship of acuity to the four conditions of intensity for the Ives grating and the artificial pupil

one-percent level in all four experiments. *Positions*, however, is significant for the grating but not for the broken ring. Significance for this variable would indicate that astigmatism is the rule rather than the exception. These results would indicate that the broken ring is a less effective test object for detecting astigmatism than the grating. The primary variable *eyes*, when evaluated by the interaction S × E, is not significant in any of the four experiments. This would indicate that neither right nor left eye had a consistently higher acuity in the majority of the 10 *S*s.

Many of the interactions are significant in one or more of the four experiments. They are difficult to interpret, however, both statistically and in terms of their visual significance. It is noteworthy that all first order interactions involving *subjects* are significant in all four experiments, whereas no first order interaction that does not involve *subjects* is significant in all four experiments.

The interactions are not of prime importance in evaluating the experiments. Where an interaction is significant, such as S × I, it need only be said that the *S*s reacted differently to the different intensities. Where an interaction is not significant, such as I × P (at the one-percent level, Table II), it may be said that the

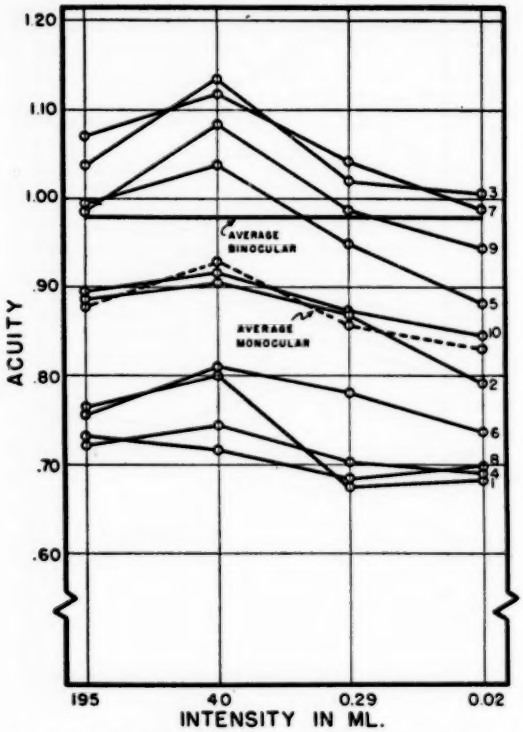


FIG. 3. The relationship of acuity to the four conditions of intensity for the Landolt broken ring and the normal pupil

form of the function relating acuity to intensity does not change significantly for the four positions.

Analysis of the intensities.—The variable *intensity* represents one of the original hypotheses of this study. It was hypothesized that it should account for some of the binocular-monocular acuity difference. It may first be noted that binocular acuity for the *S*s as a group is significantly higher than monocular acuity in all four experiments (Figs. 1–4).

Reference to Figs. 1–4, showing the acuity levels achieved by the 10 *S*s under the four conditions of intensity, may make the comparison of the intensities clearer. The variable *intensity* has already been shown to be significant by the analysis of variance in all four experiments. It is now pertinent to examine the two most important conditions of intensity. The hypothesis that the best condition of intensity should be that in which it is equal in both eyes and the most unfavorable condition should occur when the intensity in the two eyes differs widely is borne out by the results. Significant differences exist at or beyond the one-percent level between the acuities for the 40 ml. intensity and for the 0.02 ml. intensity. The statistical significance of these differences is shown in Table III. These differences have been

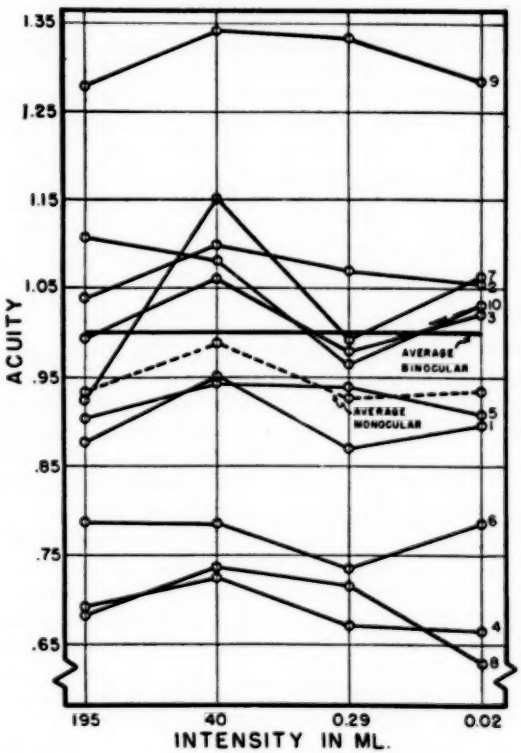


FIG. 4. The relationship of acuity to the four conditions of intensity for the Landolt broken ring and the artificial pupil

evaluated by the *t* technique for small samples using the interaction variance appropriate to the difference evaluated (S × I).[6] It may be concluded that brilliance differences play a significant part in explaining the difference between binocular and monocular visual acuity.

It should be noted that under certain conditions the two primary variables, *intensity* and *pupil changes*, are not entirely independent. This is true in the first and third experiments in which the normal pupil is employed. When intensity is varied, there are, of course, accompanying pupil changes. That intensity qua intensity is important, however, is shown by the results of the second and fourth experiments where the artificial pupil is employed to eliminate pupil variation.

The role of the pupil.—The effect of the artificial pupil may be evaluated first by an analysis of the monocular judgments obtained with and

[6] These differences may be evaluated in the usual manner, using the standard error of the distribution of differences to evaluate the mean differences. These two methods are entirely comparable and give identical results, differing only by chance fluctuations. All differences fall approximately at or beyond the one-percent level of significance when evaluated in this manner.

TABLE III

SIGNIFICANCE OF DIFFERENCES BETWEEN ACUITY LEVELS FOR 40-ML. INTENSITY AND 0.02-ML. INTENSITY

	Ives Grating Normal Pupil	Ives Grating 2-mm. Pupil	Broken Ring Normal Pupil	Broken Ring 2-mm. Pupil
N	10	10	10	10
$M_{diff.}$	0.0838	0.0709	0.09995	0.05410
S × I variance	0.03906	0.03200	0.02518	0.05927
$s_{diff.}$	0.01397	0.01265	0.01122	0.01721
t	6.00	5.60	8.91	3.15
5% level	2.26	2.26	2.26	2.26
1% level	3.25	3.25	3.25	3.25

without the artificial pupil. The results for the grating show that the two-mm. pupil tended to reduce the monocular acuity. Eight out of ten

TABLE IV

STATISTICAL EVALUATION OF CHANGES IN MEAN MONOCULAR ACUITY DUE TO THE ARTIFICIAL PUPIL

	Ives Grating	Broken Ring
N	10	10
$M_{diff.}$	0.11004	− 0.06977
$sM_{diff.}$	0.039297	0.040444
t	2.80	1.73
5% level	2.26	2.26
1% level	3.25	3.25
N		10
Mean difference of the paired differences		0.18821
$sM_{diff.}$		0.04043
t		4.66
5% level		2.26
1% level		3.25

Ss achieved lower mean monocular acuities[7] under the artificial pupil, and two *Ss* achieved higher mean monocular acuities. When the paired mean differences for each *S* between the acuities obtained with and without the artificial pupil are evaluated statistically, the probability is .02 that this reduction could have occurred by chance (Table IV). These results are reversed for the broken ring. The mean monocular acuities of five *Ss* were increased and five were decreased under the artificial pupil. The net effect is an increase in mean monocular acuity. The probability that this increase could have occurred by chance is .12.

This reversed effect for the two test objects has been evaluated statistically in the lower half of Table IV. Here the differences between the paired differences yield a probability beyond the one-percent level. It may be concluded that the use of the

TABLE V

STATISTICAL EVALUATION OF THE DECREASE IN BINOCULAR-MONOCULAR DIFFERENCES DUE TO THE ARTIFICIAL PUPIL FOR THE IVES GRATING

	Normal Pupil	Artificial Pupil (2 mm.)
N	10	10
Mean binocular-monocular diff.	0.13164	0.08060
$sM_{diff.}$	0.0166	0.01343
t	7.93	6.00
5% level	2.26	2.26
1% level	3.25	3.25
N		10
Mean difference of the paired differences		0.05104
$sM_{diff.}$		0.01548
t		3.30
5% level		2.26
1% level		3.25

[7] All acuities are in decimal units equivalent to the Snellen fraction.

artificial pupil results in different effects for the two test objects.

The second comparison for the effect of the artificial pupil is that for binocular and monocular differences. In Figs. 1–4, the average binocular acuity may be seen as the straight line. With the normal pupil, the difference between the average binocular acuity and the average monocular acuity is large relative to the same differences with the artificial pupil. This decrease of the difference between the average binocular and monocular acuity levels, due to the use of the artificial pupil, can be evaluated statistically. This has been done (Tables V and VI) by obtaining the binocular-monocular difference for both test objects with and without the artificial pupil. The paired difference between the two paired sets of differences for normal and artificial pupil (see lower half of Tables V and VI) then yields a statistical estimate of the effect of the pupil.

TABLE VI

STATISTICAL EVALUATION OF THE DECREASE IN BINOCULAR–MONOCULAR DIFFERENCES DUE TO THE ARTIFICIAL PUPIL FOR THE LANDOLT BROKEN RING

	Normal Pupil	Artificial Pupil (2 mm.)
N	10	10
Mean binocular-monocular diff.	0.10459	0.05678
$sM_{diff.}$	0.01142	0.01600
t	9.16	3.55
5% level	2.26	2.26
1% level	3.25	3.25
N		10
Mean difference of the paired differences		0.04781
$sM_{diff.}$		0.02043
t		2.34
5% level		2.26
1% level		3.25

This decrease in the binocular-monocular difference due to the effect of the artificial pupil is significant beyond the one-percent level for the grating. The comparable significance for the broken ring is between the four- and five-percent levels. The conclusion seems justified that pupil changes, like brilliance differences, account for a significant part of the superiority of binocular visual acuity.

INTERPRETATION OF THE RESULTS

The analysis of the results showed that both pupil changes and brilliance differences play a large role in accounting for the difference between binocular and monocular visual acuity. Since pupil changes may be termed a peripheral factor, and since it is less complex than the factor of brilliance differences, it would be well to discuss these results first.

Control of the pupil.—The effect of the artificial pupil differed for the two test objects. The mean monocular acuity of the majority of the *S*s was decreased for the grating and the reduction is at the two-percent level of confidence. For the broken ring, although five *S*s increased their mean monocular acuity and five *S*s decreased, the net effect was an increase with a probability of .12 of occurring by chance. The difference between the reversed effects is significant beyond the one-percent level.

It is probable that this reflects a basic difference between the two test objects. The gap in the broken ring is evidently qualitatively as well as quantitatively different from the bars in the grating. The length of the test objects differs greatly, and it is possible also that irradiation effects are different for the two test objects. It is also known that decreasing the diameter of the pupil will decrease

spherical aberration and increase diffraction effects. It is possible that these factors affected the grating in a different manner than the broken ring.

Regardless of the different effects of the artificial pupil on the two test objects, in one respect the results are similar. The important fact is that the binocular-monocular difference is reduced, and this is found for both test objects, although the broken ring shows this effect to a more considerable extent than the grating. These main results are not out of line with what might have been expected. It may be concluded that pupil changes account for part of the observed difference between binocular and monocular visual acuity.

Visual interaction.—Visual interaction indicates the gross category under which pattern and brilliance differences may be subsumed. The term is not wholly satisfactory, but is used to include all processes initiated in the two retinae and interacting centrally. It has been demonstrated that binocular acuity is significantly greater than monocular acuity. It has also been shown that the variable *intensity* is significant, and accounts for a part of the binocular-monocular difference. In this section, possible explanations in terms of central interaction will be made to account for these facts.

1. *Rivalry.*—From the stimulus situation the individual is presented with two differing monocular fields. To the tested eye is presented the stimulus object, whereas the other eye views an occluder that presents a dark field as well as an unpatterned one. The differing fields may produce rivalry and concomitant suppression. In a narrow sense, this is simply a temporal fluctuation of the fields.

Mention of rivalry as a possible explanation of binocular superiority was made in the Session Reports of the Vienna Ophthalmological Society by Lauber (**50**). Evidence lending support to this suggestion was obtained in the present experiments. Three *S*s reported that the dark field of the occluded eye tended now and again to 'close in' on the bright field of the test object and obscure it. This occurred at unpredictable and uncontrollable times. Blinking and a renewed attention to the test object was necessary to overcome it. Since these *S*s were all naive and unaccustomed to introspective analysis in visual research, the fact that three *S*s reported the phenomenon yields excellent presumptive evidence of its occurrence.

The 40-ml. intensity was that condition in which the intensity in the occluded eye equalled that in the tested eye and removed any conflict due to brilliance differences. This condition should have predisposed to minimum rivalry and, conversely, the 0.02-ml. intensity presented a very dark field to the occluded eye and should have predisposed to maximum rivalry.

It must be stated, however, that it seems unlikely that rivalry alone could account for much of the observed binocular-monocular difference. Any rivalry that occurred was necessarily infrequent. We may deduce this from the factors involved in rivalry. First, that image usually predominates which has the higher intensity. Again, contours usually tend to dominate over plain fields. Finally, Meyer and Helmholtz have indicated a third possible reason for the paucity of rivalry to be expected in this situation. The test object would predominate over the plain field because the attention is turned to it and it tends to command interest.

Although all of these factors tend to minimize rivalry in the present experiments, it is doubtful that it was entirely eliminated. Assuming that it did occur, there appear to be two ways in which it could influence acuity. First, it could distract the *S* and create a divided attention with a lowered monocular acuity as the result. Second, if in a single presentation of the stimulus object

rivalry occurred just as the threshold was being reached, the reaction of the *S* would be delayed and hence a higher value for that run would be obtained.

2. *Fusion.*—A further possibility is that fusion of the two dissimilar fields occurs. Binocular vision is a perception in which the two monocular fields, if suitable, will be elaborated into a complete, unitary, fused image at the cortical level. Fusion is related to rivalry and derives its difference only by arbitrary distinctions. That is, those monocular fields which cannot be fused due to their dissimilarity may produce rivalry.

As presented in Duke-Elder (19), fusion and rivalry appear to be an either-or phenomenon. Regardless of whether this is so, it is necessary to examine the possible consequences of fusion. It has already been stated that a fused image results in a preception that is, in general, intermediate between the two monocular fields if the two fields differ but slightly. Nevertheless, it is difficult to define the conditions under which this will occur. It can be said that there are situations in which fusion may occur, but does not necessarily have to occur. The fact that visual acuity is a foveal activity would predispose to fusion. The two foveas form corresponding areas, par excellence, the stimulation of which can lead to fusion. Also the patterns of the test objects were not of such a complicated variety that fusion with the unpatterned field would be impossible. If fusion of the two fields did occur and an intermediate perception resulted, we might normally expect a reduction in contrast and a heightened threshold. Monocular visual acuity would, in consequence, be lowered.

On other grounds, however, there is reason to suspect that the resultant of the two monocular images will not result in so simple a picture. It is questionable, first of all, whether it is valid to speak of fusion of a patterned with an unpatterned field. Unless a reduction in contrast can be discerned, the resultant will not differ from the case of complete suppression. It is also known that when the binocular combination of dissimilar fields does occur, the results are frequently unexpected. Stereoscopic combination of two fields which differ in brilliance, for example, lead to characteristic results that cannot be described as intermediate. The presentation of white to one eye and black to the other leads to a binocular perception of lustre. If the intermediate gray is presented to both eyes, the binocular combination results in a gray that bears no resemblance to the lustre produced in the first case.

Constancy would also lead us to expect a perception different than an intermediate one between the two monocular presentations. When one eye is closed the world does not appear to us to be an intermediate gray. A piece of white paper seen with one eye scarcely darkens appreciably and it takes a trained observer to notice the difference in the brilliance of the paper seen with one eye or two.

Fechner's paradoxical experiment (29, p. 522) would seem to indicate that the usual clinical occluder used to take monocular judgments would result in a less favorable perception than if the eye were actually closed. The 0.02-ml. intensity in this experiment closely approximated the clinical situation. Without defending any explicit theory, it may be noted again that this condition yielded the lowest monocular acuity. There is also the possibility, of course, that fusion and rivalry may have alternated, first one occurring and then the other. Introspective analysis by the experimenter and other qualified observers indicate that this alternation actually did occur. The conclusion seems warranted that fusion and rivalry are closely related to the reduction of monocular visual acuity.

3. *Border phenomena.*—There is one further group of related events that have not yet been mentioned. These are border phenomena and include the relationship of pattern and contour to ground and to the contrast effects induced. We know that brightness contrast has an important biological function, and there can be little doubt that

border contrast phenomena are exceedingly important in visual acuity.

The test object or stimulus in this experiment may be subsumed under the category of 'small area.' Here the ratio of contour to area is large and border phenomena dominate the picture. This is a much more important case than it would be if the area were large and the ratio of contour to area was small, since the ability to distinguish pattern or contour is essential in visual acuity.

If both eyes view the test object, it is reasonable to assume that each eye supplements the other and that enhancement of contour would occur over the situation in which the test object was viewed monocularly. That is, we may expect that simultaneous contrast would sharpen edges and yield, consequently, a lower threshold. If the two fields were slightly dissimilar, it would be reasonable to expect some reduction in contrast. If the two fields have forms which are quite different and will not easily admit of fusion into a unitary perception of a single test object, we might expect any of the following to occur:

a. The two fields may be seen superimposed on each other.

b. More usually, one of the fields will dominate. This may be expected to occur if one eye is dominant or has a much better visual acuity than the other eye.

c. One eye may see parts of the total field at the same time that the other eye also sees parts of its field. Frequently, alternations will occur so that for a while only parts of one image may be visible and the other image suppressed, and then the reverse occurs. We may call this a regional rivalry (after von Kries).

It is apparent that the results of this experiment can make no decision as to the relative importance of the factors discussed. These factors are, however, not in the nature of speculation, but are almost certainly involved in visual acuity and contribute to the binocular-monocular difference. They form a substantial part, at least, of what in the past has been labelled cortical interaction. When coupled with pupil changes, a large portion of the superiority of binocular visual acuity is explained.

Other possible factors.—The two experimental variables of *pupil changes* and *intensity* accounted for much of the binocular-monocular difference. However, not all of this difference was explained by these two conditions. We may conjecture here as to what other factors may be involved.

1. It is possible that astigmatism can account for some of this remainder. It has been noted that binocular vision could necessarily capitalize on an optical defect, such as astigmatism if it differed in the two eyes. That is, if the astigmatism which was present in all *S*s differed in degree in the different axes, binocular acuity would be enhanced in a purely optical fashion.

2. As mentioned above, pattern differences remained in the experimental situation. Only brilliance differences were tested experimentally. It is easily possible that the pattern differences could account for all of the remaining binocular superiority.

3. It is possible that chance factors as introduced by Pirenne (**42**) and Bárány (**9**) may account for the residual. This would ascribe the cause to the belief that there is more chance of seeing the test object at the threshold with two eyes than with one if independence is assumed for both eyes.

(Manuscript received
November 19, 1948)

References

1. Adler, F. H. *Clinical physiology of the eye.* New York: The Macmillan Co., 1933.
2. Adler, F. H., & Fliegelman, M. Influence of fixation upon visual acuity. *Arch Ophthal.*, 1934, **12**, 475–483.
3. Adrian, E. D., & Matthews, R. The action of light on the eye. I. The Discharge of impulses in the optic nerve and

its relation to the electric changes in the retina. *J. Physiol.*, 1927, **63**, 378–414.

4. Adrian, E. D., & Matthews, R. II. The processes involved in retinal excitation. *J. Physiol.*, 1927, **64**, 279–301.
5. Adrian, E. D., & Matthews, R. III. The interaction of retinal neurones. *J. Physiol.*, 1928, **65**, 273–298.
6. Andersen, Emelie E., & Weymouth, F. W. Visual perception and the retinal mosaic. I. Retinal mean local sign—an explanation of the fineness of binocular perception of distance. *Amer. J. Physiol.*, 1923, **64**, 561–594.
7. Averill, H. L., & Weymouth, F. W. Visual perception and the retinal mosaic. II. The influence of eye-movements on the displacement threshold. *J. comp. Psychol.*, 1925, **5**, 1925.
8. Aubert, H. *Physiologie der Netzhaut.* Breslau: Verlag von E. Morgenstern, 1865.
9. Bárány, E. A theory of binocular visual acuity and an analysis of the variability of visual acuity. *Acta ophthal.*, 1946, **24**, 63–92.
10. Bartley, S. H. Some factors in brightness discrimination. *Psychol. Rev.*, 1939, **46**, 337–358.
11. Bartley, S. H. *Vision: a study of its basis.* New York: D. Van Nostrand Co., 1941.
12. Cobb, P. W. The influence of illumination of the eye on visual acuity. *Amer. J. Physiol.*, 1911, **29**, 76–99.
13. Cobb, P. W. The influence of pupillary diameter on visual acuity. *Amer. J. Physiol.*, 1915, **36**, 335–346.
14. Cobb, P. W. Individual variations in retinal sensitivity and their correlation with ophthalmological findings. *J. exp. Psychol.*, 1922, **5**, 227–246.
15. Cobb, P. W., & Moss, F. K. Four fundamental factors in vision. *Trans. Illum. Engng. Soc.*, 1928, **23**, 496–506.
16. Crawford, B. H. Ocular interaction in its relation to measurements of brightness threshold. *Proc. roy. Soc.*, 1937, **123**, 70–73.
17. Creed, R. S. The physiological integration of sensory processes within the gray matter of the nervous system: a critical review. *Brain*, 1931, **54**, 29–54.
18. Creed, R. S. The after-image of black. *J. Physiol.*, 1931, **73**, 247–266.
19. Duke-Elder, W. S. *Text-book of ophthalmology.* Vol. 1. St. Louis: The C. V. Mosby Company, 1934.
20. Emsley, H. H. *Visual optics.* London: Hatton Press, Ltd., 1939.
21. Ferree, C. E., Rand, G., & Buckley, D. Study of ocular functions with special reference to the lookout and signal service of the navy. *J. exp. Psychol.*, 1920, **3**, 347–356.
22. Fry, G. A. The relation of border contrast to the distinctness of vision. *Psychol. Rev.*, 1931, **38**, 542–549.
23. Granit, R. The retina as a nervous center. *Arch. Ophthal.*, 1931, **6**, 104–111.
24. Hartridge, H. Chromatic aberration and the resolving power of the eye. *J. Physiol.*, 1918, **52**, 175–246.
25. Hartridge, H. Visual acuity and the resolving power of the eye. *J. Physiol.*, 1922, **57**, 52–67.
26. Hartridge, H. The visual perception of fine detail. *Phil. Trans. Roy. Soc. London*, 1947, **232**, 519–671.
27. Hering, E. Der Raumsinn und die Bewegungen des Auges. Capitel 4. Von der Prävalenz und dem Wettstreite von Conturen, p. 380. In: L. Hermann, *Handbuch der Physiologie*, Dritter Band, Leipzig: Verlag von F. C. W. Vogel, 1879.
28. Helmholtz, H. von. *Physiological optics.* Translated by J. P. C. Southall. Vol. 1. Published by The Optical Society of America, 1924.
29. Helmholtz, H. von. *Physiological optics.* Translated by J. P. C. Southall. Vol. 3. Published by The Optical Society of America, 1924.
30. Holoday, Beverley E. Die Grossenkonstanz der Sehdinge bei Variation der innen und ausseren Wahrnehmungsbedingungen. *Arch. ges. Psychol.*, 1933, **88**, 419–486.
31. Hughes, J. W. The monaural threshold: the effect of a subliminal contralateral stimulus. *Proc. roy. Soc.*, 1938, **124**, 406–420.
32. Hughes, J. W. The monaural threshold: the effect of subliminal and audible contralateral and ipsilateral stimuli. *Proc. roy. Soc.*, 1940, **128**, 144–152.
33. Imus, H. A. Visual efficiency. *Hygeia*, April and May, 1941.
34. Ives, H. E. A visual acuity test object. *Elect. World*, 1910, **55**, 939–940.
35. Ives, H. E. An improved visual acuity test object. *J. opt. Soc. Amer.*, 1917, **1**, 101.
36. Lythgoe, R. J. The measurement of visual acuity. *Med. Res. Coun., Spec. Rep. Ser.*, No. 173, 1932.
37. Lythgoe, R. J., & Phillips, L. R. Binocular summation during dark adaptation. *J. Physiol.*, 1938, **91**, 427–436.

38. Lythgoe, R. J., & Tansley, K. The relation of the critical frequency of flicker to the adaptation of the eye. *Proc. roy. Soc.*, 1929, **105**, 60–92.
39. Marshall, W. H., & Talbot, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. *Biological Symposia*, Vol. VII, p. 117. Jaques Cattell Press, 1942.
40. Parsons, Sir J. H. The visual discrimination of two points. *R. L. O. H. Reps.*, 1914, **19**, 264–273.
41. Parsons, Sir J. H. The influence of illumination on visual acuity. *R. L. O. H. Reps.*, 1914, **19**, 274–289.
42. Pirenne, M. H. Binocular and uniocular threshold of vision. *Nature*, 1943, **152**, 698–699.
43. Reeves, P. Rate of pupillary dilation and contraction. *Psychol. Rev.*, 1918, **25**, 330–340.
44. Smith, R. *A compleat system of opticks.* Cornelius Crownfield, Cambridge, England, 1738.
45. Troland, L. T. The theory and practice of the artificial pupil. *Psychol. Rev.*, 1915, **22**, 167–176.
46. Troland, L. T. Apparent brightness; its conditions and properties. *Trans. Illum. Engng. Soc.*, 1916, **11**, 947.
47. Troland, L. T. The present status of visual science. *Bull. Nat. Res. Coun.*, 1922, **5**, 27.
48. Wagman, I. H., & Gullberg, J. E. The relationship between monochromatic light and pupil diameter. The low intensity visibility curve as measured by pupillary measurements. *Amer. J. Physiol.*, 1942, **137**, 769–778.
49. Wagman, I. H., & Nathanson, L. M. Influence of intensity of white light upon the pupil diameter of the human and of the white rabbit. *Proc. Soc. exp. Biol. Med.*, 1942, **49**, 466–470.
50. Wiener Ophthal. Gesellsch. *Klin. Monatsbl. Augenheilk.*, 1927, **78**, 703.

THE ELECTRICAL RESPONSE OF THE HUMAN RETINA DURING DARK-ADAPTATION [1]

BY E. PARKER JOHNSON

Bowdoin College

Introduction

Since without sacrificing human subjects we are prevented from obtaining records of action potentials from the optic nerve or from making chemical analyses of the human retina under various conditions, we must turn to animals for studies of this nature. But the way is open for us to investigate at least one physiological aspect of the human visual process. The electrical response of the retina upon stimulation by light may be recorded, without injury to the subject, merely by placing one electrode on the cornea and another on some nearby portion of the body where it serves to make a diffuse, indirect contact with the back of the eyeball. Because of the discomfort hitherto attendant on the application of the corneal electrode even this technique has been confined largely to animals. Nevertheless, enough records of human 'electroretinograms' have been made to establish the fact that they exist and are in many ways comparable to animal responses.

Recently some rather penetrating research has been carried out (1, 2, 3) indicating a significant degree of parallelism between electrical-retinal and sensory events in man. In this work, as in previous studies, only brief periods of recording could be employed, because each blink of the subject displaced the wick-electrode from the cornea, requiring that a light be turned on while it was readjusted. Under these conditions no protracted series of records could be obtained, particularly no such series as would be required in a study of dark-adaptation, where any illumination would interrupt the whole course of the experiment.

Fortunately a method which does away with this difficulty has become available with the development of the contact-lens. The corneal electrode, instead of being applied directly, may be mounted in a contact-glass and this glass be inserted beneath the lids like an ordinary lens. The use of such a device was first reported by Riggs in 1941 (27). This improvement makes it practicable to employ precise methods of recording and to subject the results to quantitative analysis.

A rough correspondence between aspects of the electrical response and certain sensory determinations has long been acknowledged. Dewar and McKendrick in 1873 (11) reported that the amplitude of the response to stimuli of different intensities tended to follow Weber's law. Kuhne and Steiner (25) found that light-adaptation reduced the magnitude of the response while dark-adaptation restored it. Dewar and McKendrick (9, 10) and Holmgren (19) showed that wave-lengths in the mid-spectral region, which appear most bright to the human eye, were also most effective in evoking the electrical retinal response.

By far the greater part of the work attempting to link electrical with sub-

[1] This paper constitutes the major portion of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology at Brown University, June, 1947. The writer wishes to express his sincere appreciation to Professor Lorrin A. Riggs under whose guidance the study was carried out.

jective phenomena has been directed specifically toward the most prominent single component in the electrical response, the first positive peak (positivity at the cornea) known as the *b*-wave. The electrical response is, however, a very complex phenomenon the relation of whose parts, one to another, is still not understood. A slower positive component has frequently been noted following the initial positive on-response. In spite of the fact that some such slow response may appear when only the posterior of the eye is being recorded from (7) or when the iris has been paralyzed by drugs (6, 24) it seems obvious that the iris- or accommodation-artifact has frequently been mistaken for a bona fide retinal component. Karpe (22) has demonstrated clearly the nature of this artifact. It is difficult to estimate how many of the '*c*-waves' presented in the literature may be thus accounted for. Fortunately very little point has been made of the appearance of this component in recent research. There exists also a negative component (or components) designated as the *a*-wave whose presence or absence must always be assumed in detailing the magnitude of the positive effects.

Besides the 'orthodox' *a*-, *b*-, and *c*-components, numerous others have been described from time to time. Most of these have been referred to as additional *b*-waves or complications of the *b*-wave. Gotch first noted these in 1903 (13). Attempts have been made to assign these extra waves or irregularities to rods and cones or to several types of cone (1, 2, 3, 15, 16).

The existence of two distinct *b*-wave components in the human response has been made evident by the work of Adrian (1, 2). The component having the shorter latent period is characterized as 'photopic,' the slower as 'scotopic.' Adrian's electrodes were a clip, fastened to the lower part of the cheek, and a moist thread which made contact with the eyeball on the nasal side of the cornea. The electrode was removed and replaced frequently between recordings but records obtained at different times appear comparable. A three-channel, resistance-capacity coupled amplifier and an ink recorder, of the type now regularly employed in the recording of electroencephalograms, were used. With this apparatus the record of an applied steady potential fell to half its initial value in 0.4 sec. The form of the response was, therefore, somewhat distorted, though not seriously. He used color filters of the Wratten monochromatic series (Nos. 70 through 76), exploring various stimulus areas and noting also the effect of varying intensity. Although the degree of adaptation was not precisely controlled, conditions representative of light- and dark-adaptation were employed. With the eye light-adapted, the brief 'photopic' response appeared in response to all the stimuli except the blue (Wratten 76) which elicited no response. The dark-adapted eye yielded the 'photopic' response to red (Wratten 70) and mixed (dual) responses to orange or yellow. The blue end of the spectrum elicted 'scotopic' responses.

Adrian used a series of intensities and showed that the form of the response depended on the wave-length of the stimulating light. An increase to maximum intensity served to bring in but the faintest indication of a 'scotopic' response to red, while a decrease in the intensity of the blue resulted only in a 'scotopic' response of steadily decreasing amplitude. With light of intermediate wave-lengths the proportional representation of the 'photopic' component in the response increased with the intensity of the stimulus.

The present investigation is concerned chiefly with the principal *b*-wave, which Adrian has designated the 'scotopic' component. The gradual increase in the amplitude of this portion of the response, as elicited by flashes of light of several different colors, has been recorded during the course of dark-adaptation.

Apparatus and Procedure

General methodology.—The method is similar to that employed by Hartline (17) in his study of dark-adaptation in the eye of *Limulus* and by

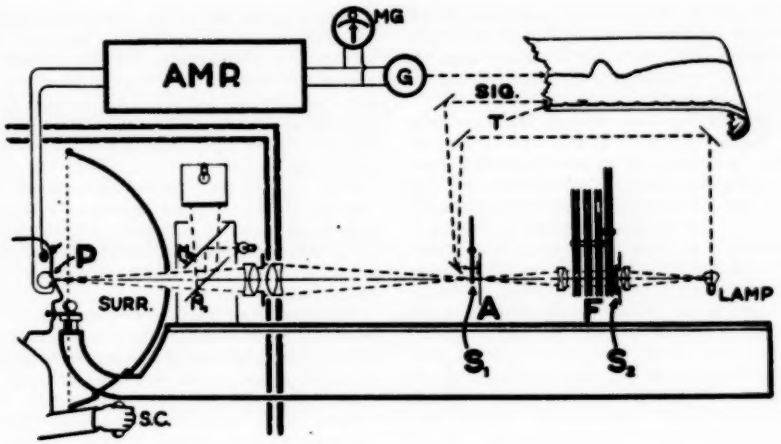


FIG. 1. Plan of the apparatus

Riggs (26) in a similar study on the frog. The eye is first dark-adapted, then light-adapted, using a standard brightness and a standard duration of adaptation. Then the adapting light is turned off and short test-flashes are presented at intervals during the course of dark-adaptation. The electrical responses to these flashes are recorded photographically on a moving film.

Optical system.—The apparatus (Fig. 1) consists of an optical system for stimulating the eye and a system for recording the electrical response. In the optical system light from a 50-CP tungsten-filament lamp is collimated and passed through a system of neutral and colored filters (F). The beam is brought to a focus at an aperture (A) having a diameter of 1.5 mm. A slotted, pendulum-type shutter (S_1) is located directly behind this aperture so as to effect the sudden exposure and extinction of the stimulus. The use of the safety-shutter (S_2) will be described later. The light is again concentrated by lenses to form an image at a point (P) which, when the subject is properly in position, falls in the plane of the pupil. This image is also 1.5 mm. in diameter and is centered within the pupil. Any necessity for an artificial pupil is thus obviated, since the iris, even when maximally contracted, cannot infringe on the stimulus-beam. The stimulus appears to the subject as a uniformly illuminated disc (the so-called 'Maxwellian view' of the lens) having a visual angle of 7°30′.

Light-adaptation is effected by means of an illuminated hemisphere (SURR). This has a matt-white interior surface and is illuminated by lamps located at its perimeter. A hole at its center admits the stimulus. During light-adaptation this hole is filled by light equivalent to that of the remainder of the field so that the entire visual field appears uniformly illuminated. This light 'plug' is introduced by an unsilvered glass mirror (M_1) which reflects light entering from the side.[2]

A small red fixation-point is similarly introduced, being mirrored at M_2 and M_1 to appear at the center of the stimulus-field. The intensity of this fixation-point is maintained by the subject at the minimum necessary to ensure adequate fixation. Central fixation has been employed throughout.

Photometric measurements.—Photometric measurements of the adapting field and stimulus-patch present an interesting problem. The light of the stimulus enters the eye as a pencil of rays passing through the center of the pupil. The apparent brightness of this patch is not reduced by constriction of the pupil. Light from the adapting field, however, enters the eye in the usual manner and the apparent brightness is, accordingly, modified by the size of the pupil. This means that if the two fields are equated in

[2] For the purposes of the present study it would not have been necessary to wed the adapting field and the stimulating system in this fashion inasmuch as there is no occasion for the simultaneous presentation of both. The apparatus has been designed, however, so as to permit the stimulus to be superimposed upon an unbroken field of standard brightness. A report of such a study has recently appeared (28).

brightness at a low level of intensity they will lose their equality when both are proportionally increased. At the higher intensity level the pupil will be smaller and the adapting field will appear less bright than the stimulus-patch.

In assigning a conventional brightness-value to the stimulus-patch at any level of intensity other than that at which actual matching determinations have been carried out one must take the pupil into account. If we make a correction for pupil diameter, the highest brightness of white light available as a stimulus is equivalent to about one and one half million footlamberts. In the experiment proper this highest brightness was not used. Lesser brightnesses were obtained by the use of Wratten neutral density filters, with about 7500 footlamberts (2.3 log units of neutral density filters) representing the highest intensity. For the colored stimuli, filters of known spectral transmission were employed.[3] These were: Farrand 1111 (red), an interference-type filter having its peak of transmission at 630 mμ; Wratten 73 (yellow), peak at 574 mμ; and Wratten 76 (blue), peak at 440 mμ.

Recording apparatus.—The subject sits within an electrically shielded cage. The position of the eye is maintained by a biting-board with a forehead rest. The corneal electrode is of the type described by Riggs (27). It has a chlorided silver disc sealed into a hole in a contact-lens so that the inner surface of disc and lens are flush with one another. The chlorided disc makes contact with the physiological saline solution between the contact-lens and the cornea. The disc is mounted far enough off center to avoid encroaching on the pupil. A light, flexible wire leads out from the contact-lens to the recording apparatus. A neutral electrode, consisting of another chlorided silver disc mounted in a hard rubber cup, is applied to the temple and filled with physiological saline solution. These electrodes function without further adjustment, except for adding solution to the neutral electrode, for several hours.

Four stages of amplification are employed. The stages are direct-coupled to ensure a linear relation between output and input. Each tube is operated within the linear portion of its Eg-Ip curve so that the wave-form of the electrical response is amplified without distortion. The output of the amplifier actuates the loop oscillograph (G, in Fig. 1) whose excursions are photographically recorded. A monitoring galvanometer (MG) enables the experimenter to adjust the balance. The camera records at a speed of 10 cm. per sec. An interval time marker (T) records each 0.1 sec. on the paper. An independent light-beam, broken simultaneously with the stimulus-beam, and by the same shutter, records the time and duration of the stimulus (SIG). The appearance of a typical record, consisting of a *b*-wave followed by a secondary rise above the base-line,[4] has been sketched in the upper right hand corner of the plan of the apparatus.

Standard conditions and procedure.—The procedure in recording a response was as follows: *E* gave *S* a 15-sec. and a 5-sec. warning of the scheduled stimulus. At the appointed time *E* opened the safety-shutter and started the camera. *S* then released the pendulum-shutter and the record was taken. Controls, with no stimulus-flash, showed that this procedure introduced no artifacts.

The scope of the experiment necessitated the combining of data from several experimental sessions. Standard conditions were, therefore, important. All work was at night. *S* dark-adapted 30 min., then light-adapted 15 min. to a field having a brightness of 1.73 footlamberts and extending nearly to the limits of the visual field. Responses were recorded at intervals of about two min. In control experiments it was found that stimulus flashes repeated with this frequency did not appear to affect the normal course of dark-adaptation. During the later stages of adaptation the gaps between stimuli were frequently longer. Time was counted in minutes from the beginning of dark-adaptation. The time intervals used were ½, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14–16 (plotted at 15), 18–22 (20), 26–36 (30), 44–56 (50), 64–76 (70), and 84–96 (90) min.

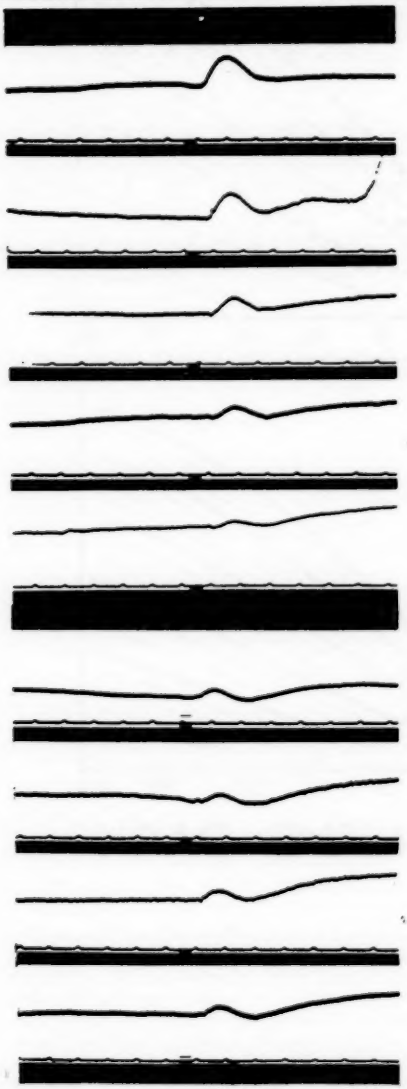
The responses recorded in this study were, with the exception of certain controls, responses to stimulus flashes having a duration of 40 milliseconds. The same subject (EPJ) served in all the reported experiments. Partial results from a second subject were confirmatory. All responses were recorded from the right eye.

Method of measuring the responses.—Slope, latency, and amplitude were considered as possible indices of the response. Amplitude was finally selected as the most feasible. Except during the first few minutes of dark-adaptation, the slope and amplitude of the *b*-wave are proportional to one another throughout the range of

[3] Peaks were determined during calibration on a Hardy recording photoelectric spectrophotometer. Data on the spectral emission of the source; photometric determinations of brightness; and transmission curves for the filters will be found in the original thesis filed in the John Hay Library of Brown University.

[4] The secondary rise is an accommodation- or iris-artifact beginning about 0.2 sec. after stimulation.

intensities employed. In the early responses, where this proportionality does not hold, the slope is relatively greater. Here the whole *b*-wave runs a briefer course in time. This means that for any given slope the maximum height reached is not so great. That responses reported as equivalent in amplitude may be unlike in other respects (e.g., latency, slope) may



0.3 *yellow* (Wratten 73 plus 0.3 neutral density filter) recorded after 86 min. of dark adaptation

0.6 *yellow*, 84 min. dark adaptation

1.0 *yellow*, 92 min. dark adaptation

1.3 *yellow*, 84 min. dark adaptation

1.6 *yellow*, 92 min. dark adaptation

3.0 *white* (3.0 neutral density filter) recorded after 6 min. of dark adaptation

0.0 *red* (Farrand 1111), 6 min. dark adaptation

0.6 *yellow*, 6 min. dark adaptation

0.6 *blue* (Wratten 76 plus 0.6 neutral density filter), 6 min. dark adaptation

FIG. 2. Sample records. The top five records represent an intensity series using yellow light, all responses recorded at about 90 minutes. The four lower records show responses of approximately equal amplitude recorded for each of the four colors after six min. of dark adaptation. The flash is indicated by the darkened area below the time line.

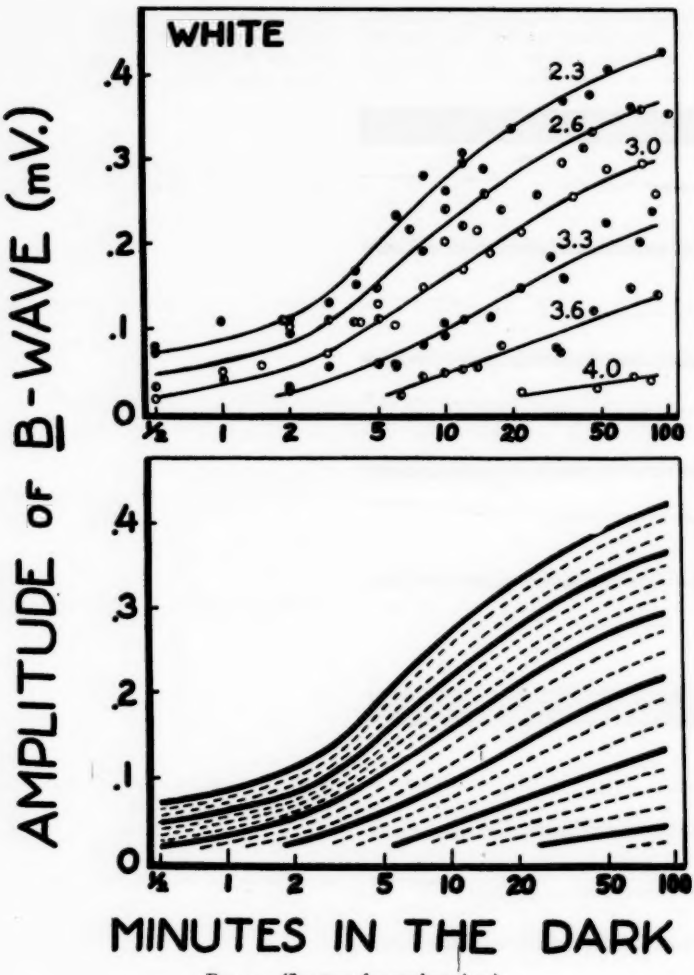


FIG. 3. (See text for explanation.)

be borne in mind later when the problem of 'threshold response' is considered.

The primary reason for choosing amplitude as the descriptive measure, rather than slope, is reliability of measurement. Records elicited by slightly greater intensities than those used in the present study show an obvious second component riding the initial slope of the principal *b*-wave. The small negative component characteristic of responses to red light is an added hazard to slope-measurement. This, when it occurs, obscures the origin and early course of the positive potential. This mingling of components occurs mostly in the early minutes of dark-adaptation. It is less pronounced later on, when the latency of the principal *b*-wave has increased. It may be noted that, even in amplitude measurement, a certain amount of arbitrariness is involved. The appearance of negative components raises a question as to what shall be taken as a base. Our own observations and those of previous investigators (29) suggest that any negative component which appears is sluggish in its return to the base-line, particu-

larly during dark-adaptation. We have, therefore, followed a practice of measuring *b*-wave amplitude from the lowest point in the record, after stimulation and preceding the positive phase of the response. In general it is only the response to red light that presents this problem and, even there, the negative potential developed is slight. Some observations based on latency, or, more properly, on the related measure known as 'implicit time,' will be presented. Latency is not, however, a very satisfactory measure of a response which develops and declines gradually. In summary, some estimate of the *b*-wave amplitude remains the most obvious and convenient measure which can be applied to the photographic records.

Results

Samples of the records are presented in Fig. 2. The data on the amplitude of the *b*-wave have been plotted in four graphs (Figs. 3 and 4). The four basic graphs show the development of the *b*-wave as elicited by each of several standard intensities of flash. A separate graph is plotted

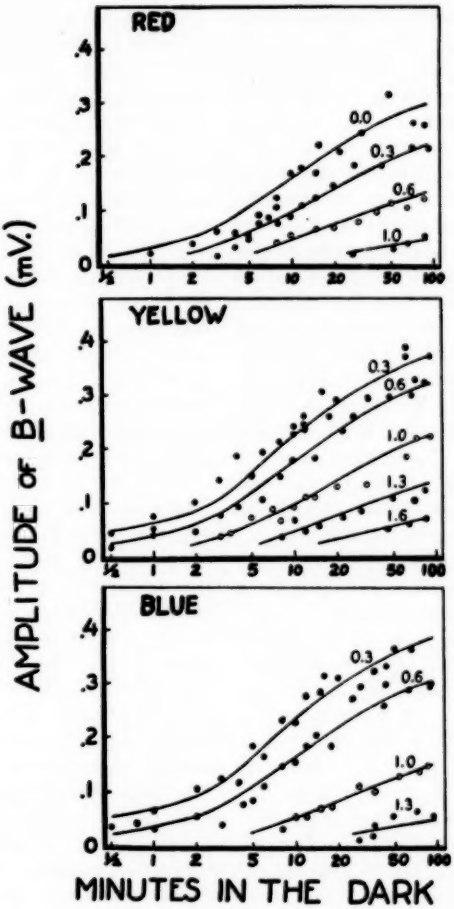


Fig. 4. (See text for explanation.)

for each of the four colors used. The response is plotted in millivolts on the ordinate. Time-in-the-dark is scaled logarithmically, in min., on the abscissa. The highest recorded amplitude of *b*-wave was 0.43 millivolts, in response to white light of the highest intensity employed (*ca.* 7500 foot-lamberts). The course of dark-adaptation was followed for nearly 100 min. in the dark. Intensity of flash is the parameter represented by the several curves appearing on each graph. Each curve is labeled with the density of the Wratten neutral-tint filter employed. Density is defined as log $1/t$, where t is the transmission of the filter.

The *white* data were plotted first and fitted, by eye, with what seemed to be a consistent set of smooth curves. In some cases the position of one curve was allowed to be determined somewhat by the positions of the neighboring curves as, for example, the early portion of the curve for 2.6 *white*. (This is the curve for responses to stimulation using 2.6 density units of Wratten No. 96 neutral-tint filters.) With these curves serving as a skeleton, a more comprehensive set was drawn up, intended to represent the amplitude of *b*-wave that might be expected in response to each 0.1-unit step from density 2.3 to density 4.1. This construction is presented below the *white* results in Fig. 3. The curves taken directly from the upper diagram are drawn in solidly while the interpolated curves are dotted.

In Fig. 4 we have the results for *red*, *yellow*, and *blue*. The 0.0 *red* stimulus (Farrand 1111 with no neutral filters) seems approximately as effective as the 3.0 *white*. If proportional changes in the intensities of white or red light have proportional effects on the amplitude of the *b*-wave we should, accordingly, expect the 0.3, 0.6, and 1.0 *red* results to fit the same curves as 3.3, 3.6, and 4.0 *white*, respectively. These are, in fact, the curves (taken from Fig. 3) drawn through the experimental points for *red* in Fig. 4. The fit seems reasonably good.

Turning to the *yellow* results we similarly find that upon equating the highest intensity used (Wratten 73 plus 0.3 neutral density) to one of the curves in Fig. 3 (2.6 *white*), we obtain a match for each proportional step in density or intensity all the way down the graph. The curves drawn through the *yellow* points are those which we have found, or might expect to find, using 2.6, 2.9, 3.3, 3.6, and 3.9 *white* as the stimulus. (Here use is made of some of the interpolated curves from the lower diagram in Fig. 3.)

We may summarize the results for for *white*, *red*, and *yellow* flashes by referring to Table I. Here, in columns (1), (2), and (3) it is seen that equivalent changes in stimulus intensity result in equivalent shifts in the effectiveness of each of these colors as an elicitor of the *b*-wave.

Now let us turn to the data for flahes of blue light as presented in column (4) of Table I. Here we find that, having made our equation of *blue* and *white* for the highest intensity of *blue* (Wratten 76 plus 0.3 neutral density), and having found this equivalent of 2.5 *white*, we must drop to 2.95 *white* to obtain a match for the next lower *blue* intensity (filter density 0.6). One would expect to obtain a match with the 2.8 *white*. The discrepancy continues to manifest itself to the same degree as denser neutral filters are used. The curves reluctantly transferred from Fig. 3 to the *blue* data at the bottom of Fig. 4 are, accordingly, those for

2.5, 2.95, 3.55, and 4.0 *white*. Observation of these values reveals that the reduction in magnitude of response occurs in log density steps exactly 1.5 times the expected size. Since the curves for *b*-amplitude as a function of time-in-the-dark are like the *white*, *red*, and *yellow* curves in other respects, the explanation which leaps first to mind is that the neutral density filters are not, in fact, 'neutral,' but filter out proportionately more of the blue light than of the other colors. The filters have been carefully calibrated, however, and while it is quite true that the transmission curves show a slight dip at the blue end of the visible range, this dip does not approach a magnitude sufficient to account for the discrepancy observed. A factor of something less than 1.2 would be indicated rather than the factor of 1.5 actually found to hold.

TABLE I

COMBINATIONS OF MONOCHROMATIC AND NEUTRAL TINT FILTERS YIELDING COMPARABLE *B*-WAVES DURING DARK-ADAPTATION

(1) Neut. density (*white*)	(2) *Red* plus density	(3) *Yellow* plus density	(4) *Blue* plus density
2.3			
2.35			
2.4			
2.45			
2.5			0.3
2.55			
2.6		0.3	
2.65			
2.7			
2.75			
2.8			
2.85			
2.9		0.6	
2.95			0.6
3.0			
3.05			
3.1			
3.15			
3.2			
3.25			
3.3	0.3	1.0	
3.35			
3.4			
3.45			
3.5			
3.55			1.0
3.6	0.6	1.3	
3.65			
3.7			
3.75			
3.8			
3.85			
3.9		1.6	
3.95			
4.0	1.0		1.3

Column (1) gives the Wratten neutral tint filter densities actually used (bold type) and the interpolated values referred to in the text. Column (2) shows the neutral densities used with the Farrand 1111 to obtain the various experimental densities of *red* light. Column (3) represents the same for *yellow* (Wratten 73). Column (4) represents the same for *blue* (Wratten 76).

Perhaps a clearer way of describing this disparity would be to say that if *b*-amplitude were plotted as a function of intensity, the curves for *white*, *red*, and *yellow* would be identical in slope for any given duration of adaptation (time-in-the-dark). They could be superimposed merely by shifting them along the intensity co-ordinate. The *blue* results, on the other hand, would differ in slope and would require the intensity scale to be expanded by a factor of 1.5 (or 1.3 if correction has been made for the lack of neutrality of the filters) to effect super-imposition.

If we confine our attention to the results for *white*, *red*, and *yellow*, we observe, among the factors of (a) stimulus intensity, (b) time-in-the-dark, and (c) amplitude of the *b*-wave, the general relationship portrayed in greatest detail by the lower set of curves in Fig. 3. To construct a 'dark-adaptation curve' for comparison with curves derived from studies of the human light-threshold we may follow a procedure justified by precedent (**17, 26**) and show the intensity required to elicit a 'constant response' at successive stages in the

course of dark-adaptation. This 'constant response' may be a specified amplitude of *b*-wave. Drawing a horizontal line across these curves at a height corresponding to our selected amplitude we may observe at what point in time each of the intensities becomes capable of eliciting this response. Plotting these intensities against their times will give us a dark-adaptation curve based on the electrical measurements. Inasmuch as the curves of Fig. 3 are not parallel, the form of the curve thus obtained will vary with the amplitude selected as a criterion. In Fig. 5 are plotted 'adaptation curves' obtained by specifying amplitudes of .05, .15, and .25 mv. Although superficially similar in appearance, these curves differ sufficiently in form to render unprofitable any discussion of 'the' electrically determined curve of dark-adaptation. It is also worth recalling that all stimuli used were far above the light-threshold in intensity.

Confirming objections to the presentation of curves of this sort is the fact that even these 'constant' responses are constant only with respect to their *b*-amplitudes. We find that .05 mv. responses recorded at one min. occur much sooner after the stimulus than do responses of the same magnitude recorded at 90 min. Since latency is as much a characteristic of the response as amplitude is, we are justified only in speaking of a constant-*amplitude* response.

Discussion and Conclusions

There are no data to be found in the conventional literature on dark-adaptation that are truly comparable with our electrical results. This is not the usual quibble over the reproduction of conditions and the differences between experimental subjects. It is the reflection of a basic difference in the type of data.

Historically, the study of adaptation is molded about the assumption that a stimulus of threshold intensity may be accepted as having a 'constant effect' upon the observer. Visual thresholds are expressed in terms of the relative amount of physical energy required to set up in the visual system the conditions necessary for the perception of this 'constant level of sensation.' In order to compare the results so obtained with any associated chemical or electrical phenomena it has been usual to suppose that the necessary conditions (chemical or electrical) in the eye would also be 'constant' at this same threshold. On this assumption, expressions of the relations between thresholds and their photochemical (**18, 21**) or electrical (**17, 26**) correlates have appeared from time to time, occasionally with reservations as to the justification of such comparisons (**26**).

In our section on results there is again presented a figure (Fig. 5) whose form is dictated by the requirements of such a comparison. A constant amplitude of the *b*-wave has been used as a criterion of sensitivity. What we see is a 'dark-adaptation curve' similar in some respects to ordinary threshold curves. No rod-cone break occurs. This is most easily explained by assuming that the principal *b*-wave, as measured, reflects only scotopic processes; but the break is often similarly absent from threshold plots when relatively low levels of preadaptation have been employed.

Such speculation assumes a relation, between our data and conventional threshold data, which may not exist; for, as has been shown, the curve or function obtained is determined to a large extent by the ampli-

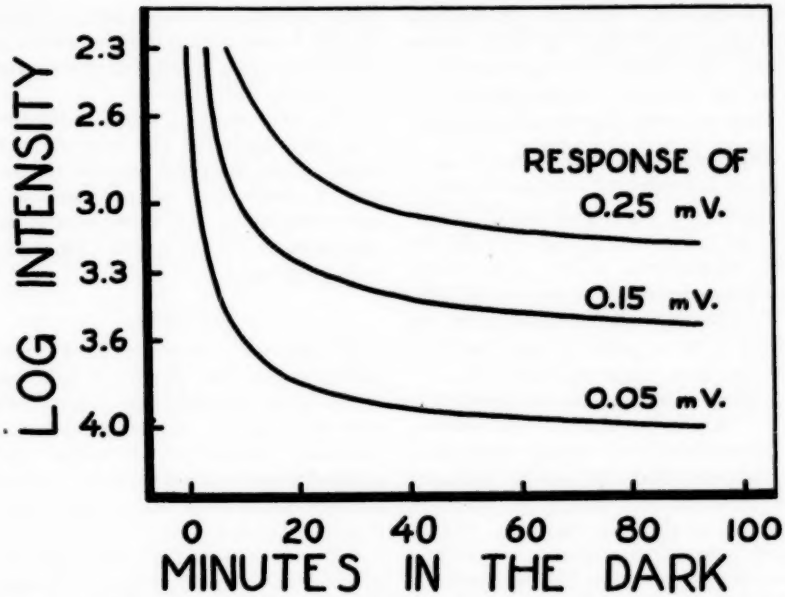


FIG. 5. 'Dark-adaptation curves' based on the *b*-wave. Three curves are drawn, representing the intensities required to elicit *b*-waves having amplitudes of 0.05, 0.15, and 0.25 mv. It may be noted that the standard amplitude selected affects the form as well as the position of the curve. The intensity scale, with its uneven intervals, is retained from Fig. 3, whence these curves are derived.

tude of the response chosen as 'standard.' If electrical responses to stimuli near the subjective visual threshold could be measured, the results might be very different. Even if such records could be obtained, however, our observations on time-factors in the responses cast serious doubt on the hypothesis that electrical events would then be found constant except, possibly, in certain limited respects—certainly not in respect *both* to amplitude and to latency. From our data we were able to determine roughtly the way in which implicit time (time from onset of stimulus to peak of response) varies with adaptation and with intensity. At any time during dark-adaptation the response may be both speeded and strengthened by the use of a higher stimulus intensity. This has been noted frequently in animals (**4, 5, 12, 20**). Dark-adaptation likewise results in an increasing amplitude of response, but is accompanied by a *lengthening* of implicit time. Thus, responses recorded at different stages of dark-adaptation, if equated for amplitude, are *not* equated with respect to time relationships. Actually there seems to be no way in which *any* intensity of light presented to the 90-min.-adapted eye can be made to produce a response similar in all respects to one obtained early in the course of dark-adaptation.

Our data might better be compared with results from psychophysical experiments carried out using similar supra-threshold levels of stimulus intensity. As these data do not exist the comparisons cannot be made.

Psychophysical studies employing conditions equivalent to those under which the electrical responses have been recorded are contemplated.

To summarize the electrical results:

The *b*-wave recorded in response to a given intensity of stimulus shows a characteristic sigmoidal rise in amplitude as a function of log time during dark-adaptation. This has been noted previously in animals (**17, 26**).

The *b*-amplitude increases with log intensity in a nearly linear fashion. The absolute rate of increase is greater the more complete the dark-adaptation. Upon looking at the earlier work on animals it seems reasonable to suppose that our study explores only the middle range of what is actually a sigmoid function. Scattered observations over a somewhat greater range confirm this supposition.

Of interest is the discovery that the principal *b*-wave, the slower of Adrian's two positive components, increases with time-in-the-dark in the same way, and at the same rate, for all colors of light—provided the intensities are adjusted to yield equipotential responses at some one time during the course of dark-adaptation. The failure of the red points to fit perfectly the earlier portion of the curve (Fig. 4) is more likely a reflection of the difficulty in measuring the response than of any basic difference in the rate of development of the *b*-wave. Significantly, the response to red continues increasing to the very end of the dark-adaptation period employed in the experiment (90 min.) just as do the responses to white, yellow, and blue.

The appearance of a small positive wave preceding the principal *b*-wave confirms the similar observation of Adrian. The apparently minor role played by the photopic mechanism, while offering a discouraging outlook for the investigation of color vision by this technique, facilitates observation of scotopic phenomena.

The necessity for using a different intensity scale for the blue data to make them conform to the pattern set by the white, red, and yellow is not explained. An experimental attempt at the resolution of this puzzle is in progress.

Summary

The electrical response of the human retina to brief test flashes of light was recorded at selected intervals of time during the course of dark-adaptation. Various intensities of white and colored light were employed in stimulation. Dark-adaptation was prolonged to ninety minutes.

All records were from the right eye of one color-normal subject. Central fixation was employed throughout. The stimulus-patch had a diameter of 7° 30′. One electrode was at the front of the eye (mounted in a contact-lens) and the other was on the temple. White, red, yellow, and blue stimuli were used. The amplified responses actuated a loop oscillograph and were recorded photographically on a moving tape. Pre-adaptation was to a field of white light having a brightness of 1.73 footlamberts.

Results are presented showing the increase in the amplitude of the principal *b*-wave as a function of time in the dark and as a function of intensity. Similarities between the responses suggest that the principal *b*-wave represents scotopic function, whatever the color of the stimulus. Red and blue responses were easily distinguished on the basis of differences in the early portions of the response.

The possibility of relating these results to our knowledge of subjective

phenomena of human vision is discussed.

(Manuscript received October 22, 1948)

References

1. Adrian, E. D. Rod and cone responses in the human eye. *Nature*, 1944, **154**, 361–362.
2. Adrian, E. D. The electrical response of the human eye. *J. Physiol.*, 1945, **104**, 84–104.
3. Adrian, E. D. Rod and cone components in the electric response of the eye. *J. Physiol.*, 1946, **105**, 24–37.
4. Adrian, E. D., & Matthews, R. The action of light on the eye: I. Discharge of impulses in the optic nerve and its relation to the electrical changes in the retina. *J. Physiol.*, 1927, **63**, 378–414.
5. Adrian, E. D., & Matthews, R. The action of light on the eye: II. Processes involved in retinal excitation. *J. Physiol.*, 1927, **64**, 279–301.
6. Brossa, A., & Kohlrausch, A. Die Aktionsströme der Netzhaut bei Reizung mit homogenen Lichtern. *Arch. f. Physiol.* (Leipzig), 1913, 449–492.
7. Chaffee, E. L., Bovie, W. T., & Hampson, A. The electrical response of the retina under stimulation by light. *J. opt. Soc. Amer.*, 1923, **7**, 1–44.
8. DeHaas, H. K. Lichtprikkels en retinastroomen in hun quantitatief verband. *Inaug. Dissert. med. Fakult. Leiden*, 1903. (Cited in **12**, **23**.)
9. Dewar, J., & McKendrick, J. G. On the physiological action of light. *Trans. R. Soc. Edin.*, 1872–1873, **27**, 141. (Cited in **12**.)
10. Dewar, J., & McKendrick, J. G. Recent researches on the physiological action of light. *Nature*, 1873, **8**, 204–205.
11. Dewar, J., & McKendrick, J. G. On the physiological action of light, No. 2. *J. anat. Physiol.*, 1873, **7**, 275–278.
12. Einthoven, W., & Jolly, W. A. The form and magnitude of the electrical response of the eye to stimulation by light at various intensities. *Quart. J. exp. Physiol.*, 1908, **1**, 373–416.
13. Gotch, F. The time relations of the photoelectric changes in the eyeball of the frog. *J. Physiol.*, 1903, **29**, 388–410.
14. Gotch, F. The time relations of the photoelectric changes produced in the eyeball of the frog by means of colored light. *J. Physiol.*, 1904, **31**, 1–29.
15. Granit, R., Munsterhjelm, A., & Zewi, M. The relation between the concentration of visual purple and retinal sensitivity to light during dark adaptation. *J. Physiol.*, 1939, **96**, 31–44.
16. Gröppel, F., Haass, F., & Kohlrausch, A. Aktionsströme und Gesichtsempfindungen des menschlichen Auges. *Z. Sinnesphysiol.*, 1938, **67**, 207–226.
17. Hartline, H. K. The dark adaptation of the eye of Limulus, as manifested by its electrical response to illumination. *J. gen. Physiol.*, 1930, **13**, 379–389.
18. Hecht, S. Vision: II. The nature of the photoreceptor process. In: C. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester, Mass.: Clark Univ. Press, 1934. Pp. 704–828.
19. Holmgren, F. *Unters. physiol. Inst. Heidelberg*, 1880, **3**, 322. (Cited in **23**.)
20. Ishihara, M. Versuch einer Deutung der photoelektrischen Schwankungen am Froschauge. *Pflüg. Arch. ges. Physiol.*, 1906, **114**, 569–618.
21. Jahn, T. L. Visual dark-adaptation. *J. opt. Soc. Amer.*, 1946, **36**, 659–665.
22. Karpe, G. The basis of clinical electroretinography. *Acta ophthal.*, Suppl. 24. Stockholm, 1948, pp. 118.
23. Kohlrausch, A. Elektrische Erscheinungen am Auge. *Handb. norm. path. Physiol.*, 1931, **12/2**, 1393–1496.
24. Kohlrausch, A., & Brossa, A. Die photoelektrische Reaktion der Tag- und Nachtvogelnetzhaut auf Licht verschiedener Wellenlänge. *Arch. f. Physiol.*, 1914, 421–431.
25. Kühne, W., & Steiner, J. Über elektrische Vorgänge im Sehorgan. *Unters. physiol. Inst. Heidelberg*, 1880, **3**, 327 and 1881, **4**, 564. (Cited in **12**.)
26. Riggs, L. A. Dark adaptation in the frog eye as determined by the electrical response of the retina. *J. cell. comp. Physiol.*, 1937, **9**, 491–510.
27. Riggs, L. A. Continuous and reproducible records of the electrical activity of the human retina. *Proc. Soc. exp. Biol., N. Y.*, 1941, **48**, 204–207.
28. Riggs, L. A., & Johnson, E. P. Electrical responses of the human retina. *J. exp. Psychol.*, 1949, **39**, 415–424.
29. Wright, W. D., & Granit, R. On the correlation of some sensory and physiological phenomena of vision. *Brit. J. Ophthal.*, Mon. Suppl. IX, London, 1938. Pp. 80.

EFFECTS OF BRIEF FLASHES OF LIGHT UPON THE COURSE OF DARK ADAPTATION [1]

BY DAVID A. GRANT AND FREDERICK A. MOTE

University of Wisconsin

The present investigation was concerned with the effect of series of short flashes of light upon the course of dark adaptation. This problem has especial significance in the interpretation of the results of a number of eyelid conditioning experiments (e.g., **2, 9, 10, 11, 12**) which were conducted under conditions which allowed the *S*s to dark adapt. It has been shown that under such conditions a secondary eyelid reaction, the β-response, increases in frequency and magnitude, which results in confusion of it and the conditioned response (**3, 4**). Previous exploration of the conditions under which this secondary eyelid response is obtained points to the dark adaptation of the rods as the essential factor in the sensitization of the β-response when the *S* sits in darkness (**5, 7**). In the present study the chief emphasis was directed toward studying the influence of series of short flashes of light upon the course of dark adaptation under conditions approximating those under which the eyelid response experiments were conducted.

No systematic investigations of the influence of flashes of light upon the course of dark adaptation have been carried out, although the results of such studies would be useful in evaluating different aspects of the retinal processes. Another principal purpose of the experiment was, therefore, to check and extend the results of the few previous investigations of the effects of flashes (**1, 14**), and to establish techniques and experimental designs for a broader and more systematic exploration of the effects of light stimulation during the course of dark adaptation.

[1] This research was supported by the Research Committee of the Graduate School of the University of Wisconsin from funds provided by the State Legislature for 1946–48.

Apparatus and Procedure

Apparatus.—The apparatus consisted primarily of a monocular Hecht-Shlaer adaptometer (**8**), an electronic timer, and a stop-watch. The adaptometer was mounted with the eyepiece projecting into a light-tight cubicle within the laboratory, so that the *S* could be dark adapted regardless of the laboratory lighting. The timer was incorporated into the pre-adaptation system of the adaptometer in such a way that the *experimental flashes* were presented to the *S* through the 35° preadapting field with central fixation. The brightness of these flashes was controlled by Wratten neutral tint fixed filters. The *test flashes* to obtain thresholds were presented on the centrally fixated 10° test field of the instrument. A camera shutter built into the adaptometer gave a constant test flash duration of 0.2 sec.

Procedure.—The general procedure followed in each dark adaptation session is outlined below:

Procedure	Time Initiated
Two min. of light adaptation	0:00
Initiate dark adaptation	2:00
Threshold No. 1 (T_1)	2:30
Threshold No. 2 (T_2)	3:00
Begin 1st Experimental Flash Series	3:00
Resume dark adaptation	6:00
Threshold No. 3 (T_3)	6:30
Threshold No. 4 (T_4)	7:00
Threshold No. 5 (T_5)	10:30
Begin 2nd Experimental Flash Series	11:00
Resume dark adaptation	14:00
Threshold No. 6 (T_6)	14:30
Threshold No. 7 (T_7)	15:00
Threshold No. 8 (T_8)	18:30
Begin 3rd Experimental Flash Series	19:00
Resume dark adaptation	22:00
Threshold No. 9 (T_9)	22:30
Threshold No. 10 (T_{10})	23:00
Threshold No. 11 (T_{11})	26:30

Procedure	Time Initiated
Begin 4th Experimental Flash Series	27:00
Resume dark adaptation	30:00
Threshold No. 12 (T_{12})	30:30
Threshold No. 13 (T_{13})	31:00
Threshold No. 14 (T_{14})	34:30
Begin 5th Experimental Flash Series	35:00
Resume dark adaptation	38:00
Threshold No. 15 (T_{15})	38:30
Threshold No. 16 (T_{16})	39:00

The initial light adaptation was established by a two-min. monocular exposure to a brightness of 1600 millilamberts on the 35° pre-adapting field of the adaptometer. Thresholds were obtained in the usual manner with this instrument (**8**), by presenting a series of test flashes and attempting to bracket the threshold. One *E* always approached the thresholds from above; the other *E* approached the thresholds from below. The experimental flashes varied in intensity and duration in the different procedures described below, under *Experimental Design*. As the general procedure outline shows, five experimental flash series were presented over the period of 37 min. Each series consisted of five flashes approximately evenly spaced during the three-min. flash series period.

During all periods of dark adaptation and between flashes during the experimental flash series the pre-adaptation field of the instrument was maintained at a background brightness of .01 millilambert. During the procedure of taking threshold measurements, the background of the adaptometer was dark between test flashes.

Experimental design.—Two brightnesses and two durations of experimental flashes were used. The brightnesses were 1600 millilamberts and 160 millilamberts, and the durations were one sec. and 0.1 sec. All four combinations of brightness and duration were used in four different sessions so that these two variables were orthogonal in the design. In addition a control procedure was used in which no experimental flashes were given. This procedure follows the general procedure, but the periods devoted to experimental flashes were given over to continued dark adaptation.

Each *S* was run through all five procedures. The order through which the procedures were run was balanced by means of latin squares so that it was possible to isolate effects due to individual variation among *S*s and ordinal effects such as practice and transfer, as well as the differences due to experimental and control procedures (**6**).

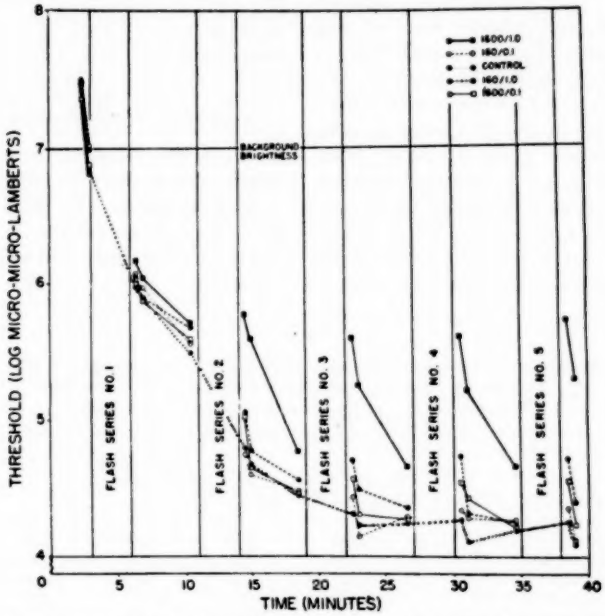


FIG. 1. The course of dark adaptation under the experimental and control conditions. Threshold (in log $\mu\mu$l) is plotted against time (in min.) for each procedure.

Subjects.—Ten *S*s were used. They were all graduate students and staff members in the University of Wisconsin Psychology Department. Each *S* had at least one practice session before going through the experiment proper. Each *E* ran five of the *S*s. A five by five latin square was used by each *E* to schedule the 25 sessions with his *S*s. For each *S* the five sessions took place on five different days which were spaced irregularly over a period of two to five weeks.

Results

All brightness measurements were converted into log micro-micro-lamberts (log $\mu\mu$l) to conform with the contemporary practice (8, p. 273).

The general results of the present experiment are given in Fig. 1, where the average course of dark adaptation is presented graphically for each of the four experimental procedures and the control procedure. Average threshold in log $\mu\mu$l is plotted against time from the initiation of light adaptation. The background brightness of .01 millilambert or 7.00 log $\mu\mu$l is indicated as a horizontal line.

The curves of Fig. 1 start in the neighborhood of 7.5 log units and drop two log units in the first 10 min. All curves, except that for the 1600 millilambert-1.0 sec. (1600/1.0) flash procedure, then drop gradually about one-half a log unit more, with slight rises following each series of experimental flashes. The 1600/1.0 procedure typically results in a rise of three-quarters to a full log unit during the third, fourth, and fifth series of experimental flashes. The data conform to the usual dark adaptation curves except for differences which may be attributed to the background brightness and to the experimental flashes.

A number of differences between procedures appear in Fig. 1. The dominant feature is the great threshold rise that is brought about by the 1600/1.0 flashes. The immediate recovery (30 sec. to 60 sec. after the flash series) also varies from procedure to procedure as does the four-minute recovery. Finally the lowest general level attained by the thresholds varies about one-half of a log unit from procedure to procedure.

Quantitative measures of these differences were obtained and analyzed. The measures were as follows:

$$\text{Threshold rise} = \tfrac{1}{3}(T_9+T_{12}+T_{13}-T_8-T_{11}-T_{14})$$

$$\text{Immediate recovery} = \tfrac{1}{3}(T_8+T_{12}+T_{13}-T_{10}-T_{13}-T_{16})$$

$$\text{Four-minute recovery} = \tfrac{1}{3}(T_9+T_{12}-T_{11}-T_{14})$$

$$\text{General level of adaptation} = \tfrac{1}{3}(T_{11}+T_{14})$$

In the above computations threshold values rather late in the course of dark adaptation (T_8 and T_9) were used in order to avoid the cone portion of the process and the rod-cone transition.

TABLE I

Average Scores for Experimental and Control Procedures on: (1) Threshold Rise, (2) Immediate Recovery, (3) Four-Min. Recovery, and (4) General Level of Adaptation

Score	Procedure				
	1600/1.0	1600/0.1	160/1.0	160/0.1	Control
1. Threshold rise	.940	.236	.329	.029	−.018
2. Immediate recovery	.386	.232	.322	.188	.133
3. Four-min. recovery	.941	.313	.415	.106	.073
4. General level of adaptation	4.446	4.229	4.291	4.264	4.201
5. Rise minus immediate recovery	.555	.003	.007	−.159	−.151
6. Rise minus four-min. recovery	−.001	−.077	−.086	−.077	−.090

The *threshold rise* score was thus the average rise in threshold brought about by the 3rd, 4th and 5th flash series. The *immediate recovery* score was the average recovery between 30 sec. and 60 sec. after the 3rd, 4th, and 5th flash series. The *four-min. recovery* score was the average recovery between 30 sec. and four and one-half min. after the 3rd and 4th flash series. The *general level of adaptation* score was the average of the thresholds taken just before the 4th and 5th flash series.

Table I contains the averaged scores for the experimental and control procedures. In general these averages show that the 1600/1.0 flashes result in greater rise in threshold with subsequent greater immediate and four-min. recovery. The immediate recovery is sufficient to overcome the rise in thresholds in all but the 1600/1.0 procedure. The four-min. recovery just overcomes the threshold rise in the 1600/1.0 procedure, but in the other procedures this length of time permits a degree of recovery greater than the rise in threshold caused by the flashes. The general level of adaptation is clearly a function of the total illumination from the experimental flashes, with some evidence for a reciprocity effect shown by the similarity of the values for 1600/0.1 and 160/1.0. That the above differences were all statistically significant was shown by the statistical analysis which is described below.

The statistical analysis of the data was moderately involved because of the experimental design which permitted elimination of the outstanding irrelevant variance. On each dark adaptation score the 25 measures for each *E*'s five *S*s were first analyzed as a single latin square (**6, 13**) according to the following breakdown of degrees of freedom:

Source	*df*
Individual Differences Between *S*s	4
Order of Procedures	4
Procedures	4
Error Estimate	12
Total	24

The means for the four experimental procedures were then analyzed according to the following orthogonal breakdown:

Source	*df*
Brightness	1
Duration	1
Brightness × Duration Interaction	1
Total	3

A pooled error estimate was obtained for each score by averaging the error mean-squares obtained by the two *E*s, and a complete interexperimenter analysis of variance was made of each of the six measures given in Table I. The summaries of these six analyses are presented in Table II. The entries in Table II are the *F*s from each analysis of variance. The bottom row contains the error mean squares upon which the *F*s of each column were based.

The *F*s in column 1 of Table II show that both brightness and duration of the flashes played a significant role in the temporary rise in threshold which interrupted the course of dark adaptation. The brightness-duration interaction was also significant, but the interpretation of this interaction is complicated,[2] as is also the case with the significant brightness-duration-experimenter interaction.

The immediate recovery was affected significantly by the duration of the flashes but not by their intensity

[2] Interaction can arise here from a number of sources. For example, if threshold effects were not linear with respect to both the brightness and to the duration of the stimuli or if there was a lack of I-t reciprocity, the interaction effect would tend to be statistically significant.

TABLE II

THE *F*-TESTS FROM THE ANALYSES OF VARIANCE OF THE SIX MEASURES DERIVED FROM THE DARK ADAPTATION RESULTS

Source of Variation	*df*	(1) Threshold Rise	(2) Immediate Recovery	(3) Four-Min. Recovery	(4) General Level of Adaptation	(5) Rise Minus Immediate Recovery	(6) Rise Minus Four-Min. Recovery
Brightness	1	111.53‡	1.81	54.98‡	18.58‡	52.98‡	3.77
Duration	1	168.13‡	12.81†	89.96‡	36.02‡	54.12‡	2.33
Experimenter	1	3.40	6.38*	0.23	0.22	0.37	4.79*
Brightness × Duration	1	27.33‡	0.06	10.50†	27.709‡	15.58‡	3.77
Brightness × Experimenter	1	0.47	0.06	0.20	1.40	1.71	0.02
Duration × Experimenter	1	0.13	0.06	0.35	0.51	3.78	4.83*
Brightness × Duration × Experimenter	1	4.47*	0.19	2.47	1.40	3.68	1.56
Error Estimate	24	*0.015*	*0.016*	*0.006*	*0.014*	*0.024*	*0.005*

* Significant at the 5% confidence level
† Significant at the 1% confidence level
‡ Significant at the 0.1% confidence level

as is shown in column 2. The difference between results of the two experimenters on immediate recovery data was significant, but may be interpreted as due to significant individual variation between *S*s (**6**, p. 437).

The four-min. recovery measures, in column 3, are significantly affected by the brightness and duration of the flashes and the brightness-duration interaction.

The relative degree of immediate recovery (rise minus immediate recovery) was significantly affected by brightness, duration, and the brightness-duration interaction as is shown by the *F*s in column 5. These findings arise because of the tremendous threshold rise in the 1600/1.0 procedure which was not compensated by immediate recovery.

The only significant *F*s in column 6, for the differences between the thresholds rise and the four-min. recovery, seem to arise because of differences between the results of the two *E*s, and these are to be attributed to variation between *S*s.

DISCUSSION

The results of the present experiment were characterized by very high reliability, as may be seen from the error variance estimates of Table II. Despite this high reliability, significant differences between the data of the two *E*s were not frequent, even though they used slightly different techniques of obtaining thresholds. Such differences between *E*s as did occur could be attributed to the individual differences between *S*s.

The fact that the brightness and the duration of the flash stimuli determine the degree of the momentary rise in threshold is in general agreement with previously reported findings (**1**, **14**); but the present findings are more consistent, which may be due to the fact that the present stimuli cover a different range than those previously reported.

That the amount of immediate recovery is affected by the duration of the light flashes but not their intensity is of especial interest, because this phenomenon has not previously been reported. Data are not available

upon which to base a complete interpretation of this observation, but an hypothesis in terms of a current theory of the photochemical processes of the retina (15) will be presented. Briefly stated, the 160 μμl and the 1600 μμl brightnesses are both well above the critical level required to convert rhodopsin to retinene and thence to vitamin A plus various protein substances; but within the temporal limits of the flashes of this experiment, duration is the critical factor in producing the irreversible retinene-vitamin A reaction. The one-sec. duration flashes converted more retinene to vitamin A than did the 0.1-sec., regardless of the intensity of the light involved. Recovery would then depend upon the relatively slow vitamin A-rhodopsin reaction. Whether this tentative interpretation holds or not will depend upon future more extensive determination of critical time factors of the photochemical reactions involved.

The fact that the general level of adaptation attained is a function of the brightness and duration of the flashes is clearly demonstrated. This finding, while not in complete agreement with some earlier results (1, 14), may be attributed to the different range of brightnesses and durations which were used in the present study.

The implications of the present findings for work on the conditioned eyelid response and on the β-response of the eyelid to light are clear. First of all, the assumption that a considerable amount of dark adaptation takes place in spite of intermittent stimulation with light flashes is justified. Interpretations that the β-response is sensitized through dark adaptation (5, 7) are thus supported. Furthermore, the actual values of the threshold changes obtained in this experiment make possible a quantitative statement of the relation between rod threshold and magnitude of the β-response. The statement, so obtained, can be the means of a critical test of the role of rod adaptation in the sensitization of the β-response of the eyelid to light.

Summary

The course of dark adaptation was studied as a function of the brightness and duration of interspersed flashes of light. The general procedure was to light-adapt the *S* and then to obtain thresholds between each of five series of five flashes of light. Two brightness, 1600 and 160 millilamberts, two durations, 1.0 sec. and 0.1 sec., and a no-flash control procedure, were used within the framework of the general procedure. In the absence of other stimulation the brightness of the adaptation field was maintained at 0.01 millilambert. Threshold measurements were obtained by means of a Hecht-Shlaer adaptometer. The thresholds were measured in log micro-micro-lamberts (log μμl). Ten *S*s were used as their own controls in a latin square design which was analyzed by means of analysis of variance so as to permit removal of variance due to practice and transfer effects, individual differences between *S*s, and differences due to *E*s. The principal findings were as follows:

1. The rise in threshold following the experimental flashes of light was a function both of the brightness and the duration of the flashes.

2. The immediate (30-sec.) recovery was great enough to compensate for the flashes in all procedures except the one with the 1600 millilambert-1.0 sec. flashes. Immediate recovery was significantly affected by the duration but not the brightness of the flashes.

3. A four-min. recovery period compensated for the flashes in the 1600 μl/1.0 sec. procedure, but the same period of time resulted in a significant drop in threshold for the other procedures over and above recovery from the flashes. This long term recovery was not significantly affected by the brightness or duration of the flashes.

4. The lowest level of threshold attained during the 37-min. adaptation period was significantly affected by the brightness and the duration of the flashes.

The findings were discussed in relation to previous work on dark adaptation and also in relation to recent observations on the conditioned eyelid response and on the β-response of the eyelid to light.

(Manuscript received
November 22, 1948)

References

1. Allen, Laura K., & Dallenbach, K. M. The effect of light-flashes during the course of dark adaptation. *Amer. J. Psychol.*, 1938, **51**, 540–548.
2. Grant, D. A. The influence of attitude on the conditioned eyelid response. *J. exp. Psychol.*, 1939, **25**, 333–346.
3. Grant, D. A. The pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1943, **32**, 139–149.
4. Grant, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.
5. Grant, D. A., & Norris, Eugenia B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.
6. Grant, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
7. Grant, D. A., Hornseth, J. P., & Hake, H. W. Sensitization of the beta-response as a function of the wavelength of the stimulus. *J. exp. Psychol.*, 1949, **39**, 195–199.
8. Hecht, S., & Shlaer, S. An adaptometer for measuring human dark adaptation. *J. opt. Soc. Amer.*, 1938, **28**, 269–275.
9. Hilgard, E. R. Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound. *Psychol. Monogr.*, 1931, **41**, No. 184.
10. Hilgard, E. R., & Campbell, A. A. The course of acquisition and extinction of conditioned eyelid responses in man. *J. exp. Psychol.*, 1936, **19**, 227–247.
11. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned responses. *J. exp. Psychol.*, 1939, **25**, 141–158.
12. Porter, J. M., Jr. The modification of conditioned eyelid responses by successive non-reinforced elicitations. *J. gen. Psychol.*, 1938, **19**, 307–323.
13. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
14. Suchman, E. A., & Weld, H. P. The effect of light-flashes during the course of dark adaptation. *Amer. J. Psychol.*, 1938, **51**, 717–726.
15. Wald, G., & Clark, Anna-Betty. Visual adaptation and chemistry of the rods. *J. gen. Physiol.*, 1937, **21**, 93–105.

EXPERIMENTAL STUDIES IN THINKING: I. SCATTERED SPEECH IN THE RESPONSES OF NORMAL SUBJECTS TO INCOMPLETE SENTENCES [1]

BY NORMAN CAMERON AND ANN MAGARET

University of Wisconsin

Language behavior and thinking, in both their genetic development and their adult characteristics, are closely interrelated and mutually interdependent. Speaking and thinking are alike organized through social interbehavior; either may be tested and validated in terms of social operations. Under special circumstances, the relationship between these two aspects of human behavior is so intimate that one may be taken as an index of the other. This is particularly true when, for example, experimental conditions call for a structured form of thinking and require that its outcome be reported in words.

Perhaps the most compelling illustrations of the interdependence of language habits and thought have come from observations made upon patients suffering from behavior disorders. In schizophrenia, for example, the cumulative restriction of social interbehavior may result in severe disorganization of both language and thinking. Impaired habits of thinking and of language behavior have not only been reported clinically; they have been demonstrated and analyzed in formal experimental situations as well (1, 2, 3, 6, 10). One of the most interesting characteristics of the disorganization found in schizophrenic communication has been identified as *scattered speech*,[2] which involves a disturbance in language continuity and, frequently, the use of terms that are imprecise approximations of the ones ordinarily employed. Verbal scattering can be especially well brought out by placing the schizophrenic patient in situations calling for active, structured thinking (1, 3, 6).

There is a basic postulate of modern behavior pathology, the *principle of continuity*, which states that any attitude or response found in patients is derived from or directly related to normal attitudes or responses (4, pp. 54 and 576). From this postulate, we should expect to find that, under properly defined experimental conditions, some of the characteristics of schizophrenic disorganization in language and thinking will appear among subjects in whom there is no reason to suspect behavior disorder. We should, for example, be able to find verbal scattering in the responses of normal individuals, and to specify the conditions under which this nonsequential or imprecise kind of response occurs. Such an extension of

[1] This research was supported in part by a grant-in-aid from the Graduate School of the University of Wisconsin.

[2] This use of the term 'scattered,' to describe discontinuity and imprecision in speech, is not to be confused with the newer and more restricted used of 'scatter' to describe intelligence test performance. Both speech and thinking have been described in the literature of behavior pathology as *scattered*, in contradistinction to *topical flight*, as far back as 1910 (12), and the term in this sense has been used continuously since then down to the present (5, 7, 8, 9, 11, 13). 'Test scatter,' on the other hand, is a more recent psychometric term, applied to the range of successes and failures a given *S* achieves on a standardized intelligence scale. Moreover, in much of the contemporary psychometric literature, the term 'intertest variability' is coming to replace 'test scatter' (15).

analysis, originally derived from behavior pathology, to the behavior of normal subjects has not yet been systematically carried out. The present paper reports the first of a series of exploratory attacks upon the general problem of identifying those conditions under which certain characteristics of language and thinking common to normal and pathological individuals may occur.

More specifically, the present investigation seeks to answer three questions:

1. Will normal subjects show *scattered speech* in their responses to situations which have been shown to call it forth in schizophrenic patients?

2. Does scattering in normal persons depend upon the particular variety of structured thinking which is demanded in the situation?

3. Is *scattered speech* related to practice, and to experimentally-induced distraction, or are there persons who are *scatter-susceptible*, that is, persons who tend to show this phenomenon regardless of temporary changes in experimental conditions?

Our results suggest that scattering is a constituent of the thinking of normal persons, that it occurs most frequently in response to situations demanding statements of causality, and that individual differences in scatter-susceptibility outweigh the effect of temporary experimentally-induced distraction or practice.

Method

Subjects.—Fifty-six women students, enrolled in the first author's introductory psychology course at the University of Wisconsin during the second semester of the academic year 1947–48, served as *Ss* in the experiment. All *Ss* were volunteers, and none knew the purpose of the experiment at the time they were tested.

Materials.—Two lists of 15 incomplete sentences each, used in an earlier study (3), were employed as stimulus material in the present investigation. These statements were originally adapted from materials used by Piaget (14) in his analysis of language and thinking in young children. One list (Q1) called for the completion of sentences involving relations of *causality*; the other (Q2) required the completion of sentences involving relations of *discordance*. The two lists of sentences follow:

Q1 (causality)

1. I'm taking psychology because . . .
2. I'll eat my (supper; lunch) because . . .
3. I had my bath because . . .
4. My hair is (brown; red; yellow) because . . .
5. A man fell down in the road because . . .
6. I am a girl because . . .
7. I go to church because . . .
8. I get warm when I run because . . .
9. A boy threw a stone at me because . . .
10. My body makes a shadow because . . .
11. The sun comes up in the morning because . . .
12. I am good because . . .
13. The wind blows because . . .
14. A fish can live in the water because . . .
15. I am alive because . . .

Q2 (discordance)

1. The glass is broken, although . . .
2. John is playing in the street, even though . . .
3. I ate another roll, although . . .
4. I have some friends, even though . . .
5. He slapped my face, although . . .
6. Mary has gone away, even though . . .
7. I have given John my bicycle, although . . .
8. It is not dark yet, even though . . .
9. I ate another piece of bread, although . . .
10. He got angry with me, even though . . .
11. It is cold today, although . . .
12. I didn't get wet yesterday, even though . . .
13. That man fell off his horse, although . . .
14. I walked for another hour, even though . . .
15. My face is white, although . . .

Two phonographically-recorded case reports provided the distracting materials.[3] Each described in some detail the development of schizophrenic disorganization in a young adult patient. One (S1) was the story of an immature adolescent girl who, in an ambiguous family situation, developed severe and somewhat bizarre schizophrenic symptoms. The other (S2) was the history of a young man whose schizophrenic

[3] The authors wish to acknowledge the assistance of Harold B. McCarty (Director) and William G. Harley, of the University radio station WHA, in recording these stories.

confusion and disorganization included unusual somatic delusions.[4]

Procedure.—Twenty-eight of the *S*s were given the two lists of incomplete sentences orally, (a) once without distraction and (b) once while listening through an earphone to one of the recorded stories. Comparisons of these *S*s' responses to the questions, with and without the distracting stimulus, constitute a unified experiment directed toward *scattering*. The other 28 *S*s (a) listened to one of the recorded stories without distraction, and (b) were given one list of incomplete sentences orally while hearing the other story through an earphone. Comparisons of these *S*s' reproductions of the stories, with and without simultaneous responses to the incomplete sentences, constitute a second unified experiment on *recall*.

The order of presentation of Q1, Q2, S1 and S2 was counterbalanced, as indicated in Table I, so that eight different procedures were employed. Seven *S*s were presented with each procedure, and the procedures were run off serially, so that *S*s numbered 1, 9, 17, 25, 33, 41 and 49 were given the first procedure, *S*s numbered 2, 10, 18, 26, 34, 42 and 50 the second procedure, and so on throughout the list.

Instructions given the *S* under condition Q1 or Q2 were as follows:

> "I am going to give you some sentences which you are to complete immediately; for example, 'An airplane flies because . . .' (or, 'It is snowing, although . . .'). Will you complete that sentence, please? All right. Whenever I give you a sentence of that sort, go right ahead and complete it as quickly and fully as you can. Do you have any questions about the procedure? All right, then here is the first sentence."

The instructions and the incomplete sentences were presented by the second author. A new question was introduced by the *E* every 45 sec., and the *S*'s responses were recorded verbatim by the first author, who was seated in the experimental room during the procedure.

Instructions under condition S1 or S2 were as follows:

> "This is an experiment in remembering. I am going to play you a recording of a story. Listen carefully to the story because when it is over I am going to ask you to write down as much of it as you can remember. Your score will be based on how much of the story you recall correctly. Do you have any questions about the procedure? All right, then listen."

Immediately after the close of the story, the *S* was taken to a private room and asked to write down as much of the story as she could remember in 15 min.

Instructions under conditions Q1S1, Q1S2 and Q2S1 were as follows:

> "This is an experiment in remembering. I am going to play you a recording of a story. Listen carefully to the story because when it is over I am going to ask you to write down as much of it as you can remember. Your score will be based on how much of the story you recall correctly. I am also going to give you some sentences which you are to complete immediately; for example, 'An airplane flies because . . .' (or, 'It is snowing although . . .'). Will you complete that sentence, please? All right. Whenever you hear a sentence of that sort, go right ahead and complete it as quickly and fully as you can, but don't forget that you must also listen to the story, because *that* is what you will be asked to recall later. Do you have any questions about the procedure? All right, then listen."

The *S*'s responses to the questions under these conditions were recorded verbatim, and at the close of the period she was given fifteen minutes to write as much of the story as she could remember.

Those *S*s who were given Q1 or Q2 as the first part of their procedure were taken to a private room for a 15-min. wait before the second part of the experiment was administered.

Results

Technique of scoring.—This experimental design permits the derivation of two sorts of quantitative score: a *recall score*, based on the number of items correctly reproduced from S1 or S2; and a *scatter score*, based on the analysis of oral responses made to Q1 and Q2.

To obtain the *recall score*, each of the stories was divided arbitrarily into meaningful units. S1 yielded 54 such units, S2 yielded 60. The *S*'s recall score was the total number of units included in her written reproduction of the story.

The *scatter score* is the sum of two independent classifications of the oral responses to the incomplete sentences, one made by each *E*. Each response

[4] Verbatim accounts of these two cases, in the form in which they were recorded and used, are to be found in (4), pp. 452–456 and 475–479.

was classified as 'scattered' or 'not scattered,' according to the following definition: [5]

> "We call a verbal statement scattered when it lacks a reasonable degree of organized sequence, when it includes imprecise, substitute or approximate words or phrases (in place of more precise, definitive ones) or when it is broken by pauses which are not followed by material that continues the trend in operation before the pause."

This definition was derived from the responses made by schizophrenic patients in an earlier investigation (3). Before attempting a classification of the responses of normal *S*s in accordance with this definition, the *E*s independently classified a random sample of these schizophrenic responses and of responses made by normal adults to the same sentences (3).

A response was considered 'scattered' by the *E*s if any *one* of the three criteria of scattering included in the definition was present. The basis for classification is indicated by the following examples of scattered responses:

1. Lack of a reasonable degree of organized sequence:

a. (I am alive because . . .) "When you are . . . either you're a living . . . you're born to live, you're born alive. (Q) Well, there are . . . if you . . . when you're alive they're . . . everything is either alive or dead and people are just alive."

b. (The wind blows because . . .) "Well . . . uh. . . uh. (Q) Because God meant it to be that way, and the wind blows . . . must be . . . sets forces in the air itself makes it . . . there's something . . . uh . . . . (Q) I don't know what I mean. I can't think it out. I can't say it. (Q) There's some composition in the air that just causes the wind."

c. (I am a girl because . . .) "I inherited those characteristics, I guess. (Q) Oh they . . . in the make-up . . . in the germ plasm happened to be that certain characteristics . . . and happened those characteristics . . . girl . . . I . . . I don't know."

d. (The sun comes up in the morning because . . .) "Um . . . um . . . . (Q) Um . . . um . . . giggle . . . . (Q) Position of the earth. I mean . . . (Q) It goes on a path and just . . . um . . . no, I can't. (Q) The tilt of the earth . . . ."

e. (It is cold today, although . . .) ". . . (Q) It isn't snowing. (Q) I don't know. (Q) It's cold . . . um . . . . (Q) Well, it's winter time . . . isn't snowing."

2. Inclusion of imprecise, substitute or approximate words or phrases, in place of more precise, definitive ones (the examples in each answer are italicized):

a. (I am a girl because . . .) "I was born that way and heredity . . . not heredity . . . I really developed from the embryo, developed the characteristics of a girl. (Q) I guess in the . . . uh . . . developed from the cell . . . just *declared that tendency of a female sex.*"

b. (I'll eat my supper because . . .) "I was hungry. (Q) I hadn't had anything to eat for several hours and due to habit I always eat supper at certain times. (Q) Well . . . why I always eat? (Q) A . . . *a certain number of hours between time* and when the time comes I'm hungry and sit down and eat."

c. (I had my bath because . . .) "*It's regularly . . . regular week . . . weekly ritual* . . . . (Q) Uh . . . . (Q) Necessary, that's all. (Q) *Something you've decided to . . . required.*"

d. (A boy threw a stone at me because . . .) "Why he hadn't tried not to do otherwise . . . through his environment . . . he was just *being foolish or revenge.* (Q) . . . (Q) . . . Well, he was . . . I'd done something to him *and he . . . just automatic.*"

e. (It is not dark yet, even though . . .) "Probably it's time to be dark. The summer months are coming on so it isn't dark. (Q) . . . *Time of year changes light more.*"

3. Occurrence of pauses which are not followed by material that continues the trend in operation before the pause:

a. (I get warm when I run because . . .) "Accelerated. I mean increased circulation in your body. (Q) Well, your heart beats faster and just . . . . (Q) Blood rushes as a result . . . uh . . . ."

b. (A man fell down in the road because . . .) "He tripped or he was walking and he slipped and fell. (Q) He was walking along, I mean . . . I can't tell you any more. (Q) He was walking . . . ."

c. (My body makes a shadow because . . .) "Well, it keeps the light from going to that space beyond me . . . makes a shadow. (Q) The space is the center space and one side of you . . .

[5] This definition agrees essentially with the established usage of the term *scattering* in behavior pathology.

The direct rays from the sun, and you'd block the light. (Q) . . . ."

d. (I am a girl because . . .) "Uh . . . uh . . . because of inheritance . . . I mean . . . uh . . . dominant genes . . . in your parents. (Q) Well, there are certain dominant characteristics and recessive characteristics and . . . . (Q) No."

e. (My face is white although . . .) "Although it's . . . uh . . . . (Q) Um . . . um . . . um . . . I can't divide my time between them. (Q) Although . . . um . . . um . . . um . . . um . . . ."

4. The following answers, classified as "not scattered," fail to show any of the three characteristics which we take to define verbal scattering, as this term is traditionally used in behavior pathology:

a. (I am alive because . . .) "Because my heart is pumping the blood and the blood carries the nutrition to all parts of my body, and the organs of my body."

b. (The wind blows because . . .) "I don't know why. (Q) It's probably air currents moving . . . the differences in air pressures high and low. (Q) If there's a low pressure area and a high pressure area, the air will go from the high pressure area to the low pressure area."

c. (I am a girl because . . .) "Oh, they're certain things that made me be born a girl. (Q) Oh, I don't know what it is in the body that determines whether you'll be a girl or a boy, (Q) Something in the . . . I don't know . . . in the germs or that sort of thing."

d. (The sun comes up in the morning because . . .) "It's daylight. (Q) When the night's over it's time for the day and the sun comes up. (Q) The earth moves around the sun and at certain times it moves farther away from the sun and therefore it's dark."

e. (It is cold today although . . .) "The sun is shining. (Q) It is cold and clear and usually if it's cloudy it's a little warm. (Q) I always think that clouds act as a blanket on the sky. (Q) Keeping it warm."

f. (It is not dark yet, even though . . .) "It's five o'clock. (Q) Well, in the winter it's almost dark at five o'clock, but in the summer it's light later. (Q) I don't know the real reason for it, but the days are longer and the sun stays up longer and therefore it's light longer."

In the present study, there were 1260 statements to be classified as 'scattered' or 'not scattered.' The independent classifications made by the two *E*s agreed with each other in 1203 of these 1260 cases—an agreement of 95.4 percent. The tetrachoric correlation between *E*s in their classifications is .95. Both the percentage of agreement and the tetrachoric r are difficult to interpret, however, since the number of responses classified as 'scattered' constitute an extremely small proportion of the total responses. Agreement between the *E*s is therefore largely a function of the considerable number of responses classified as 'not scattered' by both. When these responses are omitted from the calculations, the obtained percent of agreement remains at the reasonably high level of 85.9 percent. It seems clear, therefore, that a high degree of agreement exists between the two *E*s in their classification of the responses as 'scattered' or 'not scattered.'

A given *S*'s *scatter score*, then, is the total number of responses to the 15 Q1 or 15 Q2 sentences classified as scattered by each *E*. An *S* whose responses to all Q1 or all Q2 statements were classified as scattered by both *E*s, for example, would earn a scatter score of 30. Responses were coded so that the *E*s, at the time they made their classifications, were unaware of the experimental conditions under which each of the sentences had been completed.

Statistical analysis.—Analysis of variance[6] was applied separately to the responses of each half of the total subject population, (a) to those 28 *S*s, represented by the *scatter* procedures in Table I, who completed sentences with and without the distraction of listening to a story; and (b) to those 28 *S*s, represented by the *recall* procedures in Table I, who listened to stories with and without the distraction of simultaneously hav-

[6] The writers are indebted to Arthur J. Riopelle for his technical assistance in this analysis.

TABLE I

EXPERIMENTAL DESIGN

Scatter			Recall		
Procedure	I	II	Procedure	I	II
1	Q_1 wait	Q_2S_1 write	2	S_1 write	Q_1S_2 write
3	Q_1S_1 write	Q_2 wait	4	Q_1S_1 write	S_2 write
5	Q_2 wait	Q_1S_1 write	6	S_2 write	Q_1S_1 write
7	Q_2S_1 write	Q_1 wait	8	Q_1S_2 write	S_1 write

Q_1—incomplete sentences involving causality
Q_2—incomplete sentences involving discordance
S_1—case history of schizophrenic adolescent girl
S_2—case history of schizophrenic man

ing to complete sentences. Each *S* in each half of the experiment provided two scores; for each part, therefore, there were 28 *S*s and 56 scores.

From the statistical standpoint, there are 55 degrees of freedom in the total design of this experiment. There are also in the total design a number of factors, each of which has been subjected to analysis. These factors include: (1) the difference between Q_1 and Q_2, or between S_1 and S_2; (2) the difference between the sentences accompanied by stories and the sentences completed without such distraction; (3) the difference in order within each sub-group; and (4) the differences associated with the 28 *S*s. Out of the total of 55 degrees of freedom, one is for the difference between Q_1 and Q_2 (or S_1 and S_2); one is for the difference between sentences with and without stories; one is for order; 24 are for *S*s-within-sequences; three are for between-sequences; and one is made up of confounded interactions.

There remain 24 degrees of freedom which are used for error. These 24 degrees of freedom arise from the interactions between the *S*s-within-sequences and order, along with a number of other interactions. The error estimate is thus based upon the variation remaining after the systematic subject effects, the systematic ordinal effects, the systematic differences between the two sets of sentences, the systematic differences between sentences with and without stories, and certain of their interactions have been removed. It is, therefore, a relatively pure error estimate, although it still remains quite complex due to the possible presence of higher order interactions.

Results in terms of scatter scores.[7] —In Table II are reported the results of analysis of variance applied to procedures 1, 3, 5 and 7, in which 28 *S*s completed sentences with and without the simultaneous stimulus of a recorded story. Three items in this table are significant beyond the .005 level of confidence: (1) differences in experimental procedures, (2) type of response required by the content of

[7] The assumption of homogeneity of variance was not fulfilled in all comparisons. In fact, application of the *F*-test between a number of the variances yielded a significant difference between variances. In order to determine the possible effect of heterogeneity of variance upon the obtained results, the data were reanalyzed using the appropriate *t*-test, independent of variance heterogeneity (16, p. 83). In every case, these *t*-tests supported the results as reported in Tables II and III. We may safely conclude, therefore, that heterogeneity of variance has not distorted the obtained results.

TABLE II

RESULTS OF ANALYSIS OF VARIANCE FOR SCATTER SCORES (PROCEDURES 1, 3, 5, 7)

Source	Sum of Squares	*df*	VE	*F*
Procedure	140.053	3	46.684	9.109*
Order (I–II)	1.446	1	1.446	< 1.000
Q1 vs. Q2	54.018	1	54.018	10.540*
Q vs. QS	2.161	1	2.161	< 1.000
Confounded interactions	.875	1	.875	< 1.000
Subjects within procedures	383.286	24	15.970	3.116*
Error	123.000	24	5.125	

* Significant beyond the .005 level of confidence

the incomplete sentences, and (3) individual differences among *S*s.

The first of these findings is due primarily to the large number of scattered responses which occur under experimental condition 1 (Q1–Q2S1). Fig. 1 presents graphically the scatter frequencies, in terms of the sum of responses classified as scattered by the two *E*s independently, which were obtained under each of the four experimental conditions. It is clear that *S*s who first complete sentences of causality, and then complete sentences of discordance while listening to a story, show much higher frequencies of scattered language than do *S*s who are presented with any of the other three experimental conditions. It is possible that this finding is due largely to differences among *S*s[8].

To the other two positive findings in Table II more meaning can be attached. The difference between Q1 and Q2, for example, is in the direction of greater scatter-frequency in response to Q1 sentences than in response to Q2 sentences. The mean scatter score for Q1 is 4.8; for Q2 it is 2.9. Under the conditions of this experiment, therefore, incomplete statements requiring responses of *causality* call forth significantly more verbal scattering than do statements requiring responses of *discordance*.

Of equal importance is the demonstrated significance of individual differences in this analysis. It will be noted from Table II that while *S*s within sequences show a significant consistency in frequency of scattered responses, there is no significant difference between frequency of scattering occurring when sentences are presented alone and frequency of scattering when the task of completing sentences is combined with that of listening to a story. Neither does order of presentation (I–II) yield significant results. These findings suggest that at least two temporary experimental conditions—practice in completing sentences, and instructions to listen to and recall a story—are less important in the production of verbal scattering than are the inclusion of causal relations in the incomplete sentences, and consistent individual tendencies toward scattered or non-scattered speech.

Results in terms of recall score.—Table III presents the results of analysis of variance with the 28 *S*s following procedures 2, 4, 6 and 8. It will be noted that the difference between recall score when the story is

[8] When VE (procedures)/VE (subjects within procedures) is tested, the value of *F* becomes 2.92, which is not significant. This suggests that individual differences among *S*s may account in large part for the difference between scatter-frequency under experimental condition 1 as contrasted with the other three procedures.

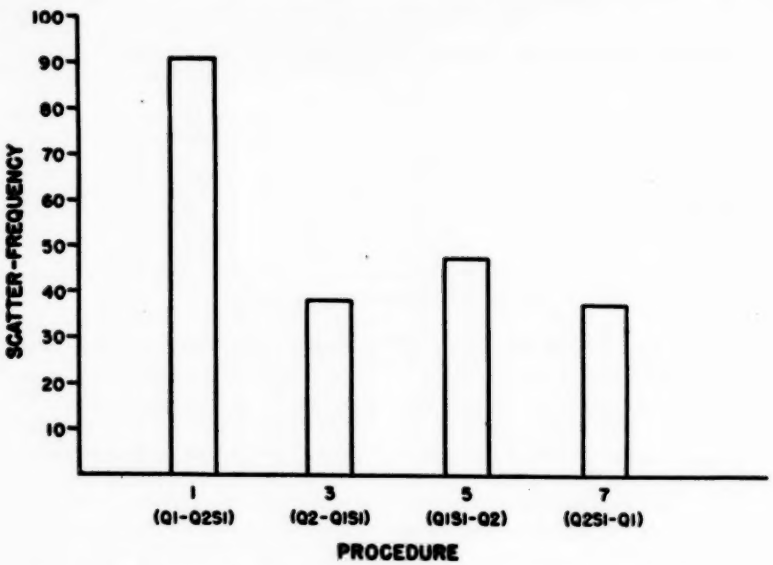


FIG. 1. Frequency of occurrence of scattered speech under four experimental conditions

heard alone, and recall score when the *S* is simultaneously engaged in orally completing sentences, is significant at the .01 level of confidence. This difference is in the expected direction: the mean recall scores for S1 and S2 heard alone are 13.64 and 12.56, respectively; the mean recall scores for S1 and S2 heard under conditions of distraction are 9.61 and 8.58, respectively. No other results of this analysis of recall score are significant, as shown in Table III. Differences in experimental procedures (2, 4, 6, 8), the order of presentation (I–II), the content of the two stories (S1–S2), and individual differences among *S*s within sequences, all fail to yield results which are statistically significant.

These findings suggest an explanation for the failure of experimentally-induced distraction to affect significantly the scatter score obtained

TABLE III

RESULTS OF ANALYSIS OF VARIANCE FOR RECALL SCORES (PROCEDURES 2, 4, 6, 8)

Source	Sum of Squares	*df*	VE	*F*
Procedures	29.928	3	9.976	< 1.000
Order (I–II)	8.643	1	8.643	< 1.000
S1 vs. S2	16.071	1	16.071	1.532
S vs. QS	216.071	1	216.071	20.602*
Confounded interactions	3.501	1	3.501	< 1.000
Subjects within procedures	305.428	24	12.726	1.213
Error	251.714	24	10.488	

* Significant beyond the .01 level of confidence

from the completed sentences, as shown in Table II. Apparently these *S*s, asked to listen to a story while simultaneously completing sentences, solved the problem of distraction by attending to the sentences and largely ignoring the story. This solution was favored by the physical presence of the two *E*s and by the concrete social situation which consequently developed. In other words, the social interbehavior of *E*s and *S* was prepotent, in stimulus value, over the phonographically-recorded material which, under simpler conditions, successfully held the *S*'s attention.

Discussion

Our findings may conveniently be discussed in the form of answers to the three questions that we posed at the beginning of this report.

1. *Normal Ss do show scattered speech in their responses to situations which have been shown to call it forth from schizophrenic patients.* No *S* was without at least one response which the *E*s classified as *scattered*; the range of frequency of scatter for these *S*s was from 1 to 27. It is clear that non-sequential, approximate and interrupted language responses of the sort originally identified in schizophrenic patients are part of the repertory of non-disordered persons. In other words, at least one characteristic of schizophrenic disorganization can be considered as derived from and related to the responses of normal individuals. It is equally clear from our findings that, under the conditions of this experiment, verbal scattering occurs with widely different frequencies among the individual *S*s.

2. *Scattering in normal Ss seems to depend upon the kind of structured thinking which is demanded by the situation.* In this investigation, it is the incomplete sentence demanding a statement of *causality* that evokes the highest frequency of scattered responses from our normal *S*s. This finding is consistent with the earlier study of schizophrenic patients (**2**, **3**), in which it was noted that schizophrenics had greatest difficulty in completing statements employing the 'because' connective. Our results, together with those reported in the study of schizophrenics, suggest that in future experimentation it may be more efficient to employ material calling for responses of causality, rather than discordance, if scattering is to be studied.

3. *Verbal scattering in this investigation is less related to practice and to temporary experimentally-induced distraction than it is to consistent individual tendencies toward scattered speech.* Our findings demonstrate no significant difference in scatter-frequency with and without temporary distraction, and no significant effect of practice. We have pointed out that this negative result may be due at least in part to the prepotency of a social situation, including the physical presence of the *E*s, over stimulus material consisting of a phonographically recorded story. Some support to this explanation is afforded by the significant decline in story recall score when story and incomplete sentences are present simultaneously.

On the other hand, our findings include significant differences among individual *S*s in frequency of verbal scattering, apart from such factors as practice and experimentally-induced distraction. Our *S*s respond in this experimental situation as if they differed in what we may call *scatter-susceptibility*—as if, in other words, certain persons have a tendency to exhibit scattered speech, whether they are being simultaneously sub-

jected to distracting stimulation or not. If we may tentatively postulate scatter-susceptibility as a variable in the thinking and language behavior of normal individuals, the obvious next step is to identify any additional characteristics of the person which may go along with this tendency.

In schizophrenic disorganization, we consider the use of non-sequential, approximate and interrupted language as one of the cumulative consequences of progressive social withdrawal. If social interbehavior has not been adequately developed, or if it becomes for any reason severely restricted and impoverished, the result may be that language and thinking are not validated in terms of social operations. Symbolic behavior may thus grow increasingly ineffectual, and eventually quite useless, as an instrument of social interaction. Viewed in this light, the scattering found in schizophrenic patients is comprehensible as an end-product of schizophrenic desocialization.

The application of such an explanation as this to scatter-susceptibility in non-schizophrenic persons, however, is no simple task. It might, of course, be argued that the experimental situation was interpreted by some of the *S*s as so artificial that it did not call forth their highest level of social communication. And it might also be held that, for some students, the unexpected presence of two faculty members as *E*s—one of them their professor—was sufficiently embarrassing and unnerving to produce language disorganization. Certainly an *S* did sometimes express surprise upon encountering the *E*s, and an occasional student showed signs that could be interpreted as indicators of embarrassment. From this point of view, scatter-susceptibility would represent a special sensitivity to the particular social elements in the experimental situation constituted by the *E*s. Such an explanation has the disadvantage that it bears a relationship to the reasons assumed to underlie scattering in schizophrenic patients which is, at most, only analogous.

A more fruitful approach to the problem of the relationship between scattering in schizophrenic patients and scattering in normal persons would lead us to take two further steps: first, the examination of other responses to situations requiring structured thinking made by the same *S*s; and second, the identification of personality variables which may be related to scatter-susceptibility. These two steps constitute the material of the next report in this series of investigations.

Summary

As the first of a series of investigations into the language and thinking of normal *S*s, scattered speech in response to incomplete sentences was studied in 56 women college students. Half of the *S*s completed sentences under two conditions: once without distraction, once while listening to a story which they were instructed to recall later. The remainder of the *S*s listened to a story under two conditions: once without distraction, once while attempting to complete orally a series of incomplete sentences. Analysis of variance applied to recall scores and scatter scores yielded the following results:

1. Significantly fewer items of the stories were recalled under conditions of distraction than without distraction.

2. Scatter score was not significantly related either to the presence of a distracting stimulus or to practice in completing sentences.

3. Significantly more scattered responses occurred when sentences involving causality were followed by the simultaneous presentation of discordant sentences and a distracting stimulus than under three other experimental conditions.

4. Sentences requiring completion in terms of causal relations produced significantly greater scatter-frequency than sentences calling for relations of discordance.

5. There are significant individual differences in scatter-frequency.

These results are tentatively interpreted in terms of the concept of *scatter-susceptibility*.

(Manuscript received January 10, 1949)

References

1. Bolles, M., & Goldstein, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–65.
2. Cameron, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 221, 1–34.
3. Cameron, N. A study of thinking in senile deterioration and schizophrenic disorganization. *Amer. J. Psychol.*, 1938, **51**, 650–665.
4. Cameron, N. *The psychology of behavior disorders: A biosocial interpretation.* Boston: Houghton Mifflin, 1947, pp. xxi + 622.
5. Diethelm, O. *Treatment in psychiatry.* New York: Macmillan, 1936, pp. 204 and 210.
6. Hanfmann, E., & Kasanin, J. Conceptual thinking in schizophrenia. Nervous and Mental Disease Monographs, 1942, pp. 1–115.
7. Henderson, D., & Gillespie, R. *A textbook of psychiatry.* London: Oxford Univ. Press, 5th ed., 1940, p. 97.
8. Hinsie, L., & Shatsky, J. *Psychiatric dictionary, with encyclopedic treatment of modern terms.* New York: Oxford Univ. Press, 1940, p. 475.
9. Hoch, A. Review of Bleuler's "Schizophrenia." New York State Hospitals Bull., August 15, 1912.
10. Kasanin, J. (Ed.) *Language and thought in schizophrenia.* Berkeley: Univ. California Press, 1944, pp. 133.
11. MacCurdy, J. *The psychology of emotion, morbid and normal.* New York: Harcourt Brace, 1925, pp. 422 and 466.
12. Meyer, A. The dynamic interpretation of dementia praecox. *Amer. J. Psychol.*, 1910, **21**, 385–403.
13. Muncie, W. *Psychiatry and psychobiology: A textbook of normal and abnormal behavior.* St. Louis: Mosby, 1948. 2nd ed., pp. 162–163.
14. Piaget, J. *Judgment and reasoning in the child.* New York: Harcourt Brace, 1928, pp. viii + 260.
15. Rabin, A. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.
16. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State College Press, 1946, 4th ed., p. 83.

VARIOUS DEGREES WITHIN A SINGLE DRIVE AS CUES FOR SPATIAL RESPONSE LEARNING IN THE WHITE RAT*

BY ROBERT BLOOMBERG

University of Tennessee

AND

WILSE B. WEBB

Washington University

INTRODUCTION

Considerable research has been stimulated by a theory of primary motivation stated by Hull in his recent book *Principles of behavior.* In this book Hull sets forth in some detail and specificity the basic hypothesis concerning the role of habit (learning) and motivational conditions (drive) in the determination of behavior.

The present study is particularly concerned with certain aspects of a postulate derived by Hull and formulated from his theory of primary motivation. This postulate is as follows (2, p. 253): "Associated with every drive (D) there is a characteristic drive stimulus (S_D) whose intensity is an increasing monotonic function of the drive in question." [1]

A particular implication developed from this postulate, and from which the present study is a departure, is Hull's Corollary XII (2, p. 251):

> Organisms will learn to react differentially to a given objective situation according to the drive active at the time, and to react differentially to a given drive according to the objective situation at the time.

This may be interpreted to mean that rather than the correctness or incorrectness of differential responses being based on external cue, e.g., black-white discrimination problems, differential responses may be developed on the basis of internal cue stimuli resulting from existing drive conditions. Empirical data concerning this corollary may be found in studies by Hull and Leeper and others. Hull's early experiment (1) made use of a two choice maze, the selection of one alley of which led to satisfaction of the drive operating at the time, e.g., hunger, while the other alley led to a door blocking the way to a goal box appropriate for a different drive, e.g., thirst. This second alley and goal box were open on other days when the second drive used in the experiment was operating while the first alley was blocked. Learning was very slow, but the animals gradually attained considerable efficiency in making the response appropriate to the drive dominant at the time.

In a study by Leeper (4) substantially similar to that of Hull's experiment, a maze similar, but differing in certain details, was used. Two distinct reinforcement chambers were used, and no passageways were blocked at any time. Thus, if on a food-motivated day the rat went to the water side, he could enter the goal chamber and find water and then retrace his steps to the food always present in the alternate goal chamber. Under this technique Leeper's animals learned to respond correctly in much less time than was necessary for Hull's subjects.

* The present work was done without knowledge of the similar work of Jenkins and Hanratty, Drive intensity discrimination in the albino rat, *J. comp. physiol. Psychol.*, and was, as a matter of fact, submitted for publication at an earlier date.

[1] For a thorough treatment of the theory of primary motivation, see Hull, C. L., *Principles of behavior*, Chapter XIV.

Kendler (3), in a study based upon the investigations of Hull and Leeper, found that it was possible for rats to learn discriminative behavior on the basis of simultaneous hunger and thirst drives during training and motivated by only one drive during test trials. In this study Kendler rewarded but one of these drives on one response and the second drive on the alternate response, under conditions of simultaneous hunger and thirst. The apparatus used was a T-maze with both alleys open at all times, but during test runs only one drive was present at any one time.

It will be noted in these studies that the cues for differential responses were the drive stimuli arising from the existing drive states. These studies used a tow-choice maze with both responses being correct at different times, but the cue for the correct response at any particular time was the internal drive state of the organism while the external environmental situation was identical for both responses. Thus, in a T-maze, the choice point offers no external cues. The correctness or incorrectness of the responses was dependent upon the internal drive state of the organism acting at the time.

All of the studies specifically mentioned above indicate the ability of the albino rat to make differential responses on the basis of differential drive states in identical external environmental situations.

The present study proposes to extend the hypothesis mentioned above to situations in which the differential drive states are various degrees of a single drive continuum. Previous studies have been alike in using two different drives either successively or simultaneously as the differential drive states. This experiment attempts to examine the possibility that various levels of intensity of a single drive may function as stimulus cues in making different responses to identical external environmental situations. During the training period the animals are forced (during early stages of training) (see training procedure) to make the appropriate response for the existing drive state and also are forced to the inappropriate goal chamber to find no reward.

On the basis of Hull's analysis (2, pp. 240–242), we should expect successful discriminative behavior on test trials if the cues arising from distinct degrees of a single drive continuum serve as discriminable drive stimuli. Thus, the stimulus complex leading to the choice of the alley appropriate for the degree of drive present would be distinct from the stimulus complex leading to the choice of the alley appropriate only for the second degree of drive.

Experiment I

A. *Experimental method*

1. *Subjects.*—Ten male and two female albino rats of approximately 100 days of age were used in the experiment. These animals were raised in the rat laboratory of the Home Economics Department of the University of Tennessee. The 12 animals were divided into two groups of six animals each for running purposes. Two of the animals died during the experiment.

2. *Apparatus.*—A single choice T-maze was used, the floor plans and dimensions of which are shown in Fig. 1.

3. *Training procedure.*

Pretraining and Position Habit Test

a. The subjects were pre-fed and handled for a period of 10 days prior to the experiment proper. Group I was fed at 4 P.M. daily, Group II at 6 P.M. daily. The animals were placed in separated feeding cages for one hour and were given nine gm. of Purina Dog Chow Biscuits. They were put back into the regular living cages following feeding and were not fed in the living cages.

b. Day Minus 2. (See Table I.) All animals were placed separately in the starting box of the maze. Here they were permitted to wander freely for one-half hour with all doors open. The animal's initial turn at the choice point was recorded.

c. Day Minus 1. All subjects were given two trials in the maze under the same conditions as on the previous day (no reward) with the exception that D-1 was down. This door was raised as the animal approached it. The two turns at the choice point were recorded for each animal. Any animal choosing the same side on the two days (total of three trials) was discarded. Three were discarded. The remaining 12 animals were randomly distributed into the two groups.

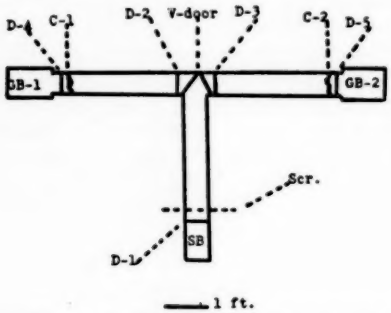


FIG. 1. Ground plan of the T-maze used in this experiment. SB, starting box, D-1, D-2, D-3, D-4, and D-5, Vertical Sliding Doors. C-1, C-2 are curtains hung before the goal boxes to prevent the animals from seeing the contents thereof when D-4 or D-5 is open. GB-1 and GB-2 are the goal boxes. The goal boxes are 12 in. long, 6½ in. wide and as deep as the alleys, which are 4½ in. and 3½ in. wide. The entire maze and the floor of GB-1 is covered with hardware cloth. The doors are constructed of Masonite and slide vertically in grooves in the alley walls. They are manipulated by means of cords which extend from a hole in the top of the door through a second hole in wooden bridges directly above these doors and 3½ in. above the alley. The strings are pulled by the experimenter standing behind the Screen, Scr. The left half of the vertical alley including the left wall, the entire left horizontal alley, GB-1, and D-2 and D-4 are painted black. C-1 is made of black cloth. The right half of the vertical alley including the wall, the entire right horizontal alley, GB-1, and D-3 and D-5 are all painted white. C-2 is made of white cloth. The entire maze was illuminated by a 100 watt bulb located approximately 4 feet above the V door. The V door shown in the figure was inserted in the maze in Experiment II and was not used in the first experiment. It is suspended and manipulated in the same manner as the other doors, but slides vertically on strips of metal nailed to the alley walls.

d. Order of trials in the training series. The animals were motivated by 3 and 22 hours of hunger during the training period and by the same drives during critical test trials. A response to the right goal box (three hours of hunger) was rewarded by .75 gm. of Purina Dog Chow Pellet and a response to the left goal box (22 hours of hunger) was rewarded by the same amount of food. Only one pellet was given on each trial, and this was contained in a metal food cup located in the rewarding goal box. The order of correct responses during the entire experiment, i.e., to the right or left, was a prearranged irregular series to prevent ordinal patterning. (See below.) In order to follow this irregular series indicated above and not to extend the training period excessively, it was necessary to reward both habits on some days, but only one or the other on some days. Each habit received the same number of reinforcements until learning was obtained. (See Table I.)

e. Procedure during the training period. Each training period consisted of five trials to the reinforcing and five trials to the non-reinforcing goal box. Responses were forced by lowering doors 2 or 4. Beginning on Day 4 of the training period, the first trial of each training series was a free trial, the second trial was a forced run to the side opposite that chosen on trial one. The last eight trials of the series followed a prearranged irregular order. Using the letter A to indicate a rewarded response, and the letter B a non-rewarded response, these trials were always given in the following order: BABABAAB.

The animals were run in groups of six and approximately five minutes elapsed between each trial.

4. *Test or critical series.*—The free trials given on the first trial of each training series were a part of the test series. The criterion for successful discrimination was set at six successive correct responses (three right and three left) on 'critical' free trials. 'Critical' trials were trials in which the reward in the previous training series had been on the opposite to the presently 'correct' side. For example, if, in a free trial, an animal correctly ran to the right goal box when motivated by three hours of hunger, this was considered a critical trial only when the reward in the last training series had been in the left goal box. (See Discussion.) It may be seen by the schedule in Table I that only one-half of the free trials were 'critical.'

B. *Results*

The criterion of learning was deliberately made a difficult one to achieve.

TABLE I

ORDER AND SIDE OF CORRECT (REWARDED) RESPONSES THROUGHOUT THE TRAINING PERIOD IN EXPERIMENTS I AND II

Day Minus 2—The day on which all animals were placed separately in the maze to familiarize them with it. The first trial for position habit was also given.

Day Minus 1—The day on which the second and third position habit tests were given.

Experiment I			Experiment II		
Day	Correct Response	Drive	Day	Correct Response	Drive
1	L	22	1	R	22
1	R	3	1	L	3
2	R	3	2	L	3
3	L	22	3	R	22
4	L	22	4	R	22
4	R	3	4	L	3
5	L	22	5	R	22
5	R	3	5	L	3
6	R	3	6	L	3
7	L	22	7	R	22
8	L	22	8	R	22
8	R	3	8	L	3
9	R	3	9	L	3
10	L	22	10	R	22
11	L	22	11	R	22
11	R	3	11	L	3
12	R	3	12	L	3
13	L	22	13	R	22
13	R	3	13	L	3
14	L	22	14	R	22
15	L	22	15	R	22
15	R	3	15	L	3
16	R	3	16	L	3
17	L	22	17	R	22

* This series was repeated in the same order until learning took place or the experiment was terminated.

It should be noted again that learning was considered to have occurred only when six (three right and three left) successive and correct responses had been made during critical trials by each animal. A critical trial was defined above as a correct choice of the right or left goal box after having been rewarded in the previous training series on the opposite side of the maze. Errors made during free trials which were not critical trials nullified preceding correct and successive critical trials unless the criterion had already been reached. (Hence 12 'correct' runs were performed to reach the learning criterion.)

Only one animal reached this criterion. (See Table II.) No other animal learned the problem and the experiment was terminated after 130 reinforced responses under each of the two motivating conditions and an equal number of non-rewarded responses. At the termination of running, five animals of the nine remaining[2] were recorded on the last critical trial as running to the side on which they had received reward on the immediately preceding training series, i.e., running incorrectly. No animal except the one which learned the problem had run as many as five correct 'critical' trials in a sequence.

[2] Two animals of the 12 had died and one animal had completed the learning task.

TABLE II

Results obtained in *both* experiments. The numbers of reinforcements prior to the criterion measure under the 3- and the 22-hour hunger drives during critical trials are given. Ten animals did not learn the problem, eight after 130 trials (Exp. I) and two after 195 trials (Exp. II).

I	II	
Animals	No. of Reinforcements to Reach Criterion	
	L	R
1	90	90
2	115	115
3	120	120
4	115	115
5	110	110
6	115	115
7	155	155
8	160	160
9	180	180
10	180	180
11	185	185

* Experiment I subject

This running to the most recently rewarded side was found to be the typical 'free trial' response. During the training sequence the 'correct' turns on the non-critical free trials, i.e., after having be rewarded on the same side as the now 'correct' side in the previous training series, never fell below 70 percent, and only in three instances fell below 80 percent. No way of eliminating this difficulty, a correct response being in opposition to the most recently rewarded experience on the critical trials, was apparent to the authors. Certainly the difficulty of learning was increased for the animal.

Experiment II

A. *Introduction*

It was believed that various factors in the original experimental design could be changed to facilitate learning. These changes included subjects, apparatus, and training procedure.

1. *Subjects.*—Twelve female albino rats of approximately 180 days of age were used. They were divided into two groups of six animals each for running purposes.

2. *Apparatus.*—A 'V' door, shown in Fig. 1, was made of hardware cloth and installed at the choice point. This door was suspended from a wooden crossbar 3½ in. above the maze and was fitted to slide vertically on strips of metal nailed to the inside of the alley walls just to the right and left of the vertical alley. It was suspended by a cord and was manipulated by the experimenter behind his screen.

3. *Pretraining and position habit tests.*—These remained the same as in Experiment I.

4. *Training procedure.*—The changes in the training procedure were as follows:

a. Groups III and IV were fed at 9 and 10 A.M. daily, respectively, throughout the experiment.

b. All animals were given as much Purina Dog Biscuit as they could eat in one hour.

c. The three-hour motivated response was rewarded in the left goal box, and the 22-hour motivated response was rewarded in the right goal box. (See Table I.)

d. Individual training series were changed to include five reinforced and two non-reinforced trials in each series. The first trial, beginning on day three of the training period, was a free trial, and thereafter, throughout the experiment, this first trial of the seven in each series continued to be a free trial. The second trial of each series was a forced trial on the side opposite that chosen on the first trial. The last five trials were forced in an irregular order prearranged separately for each series.

e. No change was made in the criterion for learning, but the V door mentioned previously was used during critical and other free trials. This door was down and remained down for five seconds after the animal reached it and then was raised. All other doors were up during free trials.

Further critical trials were given to those animals which had passed the criterion for learning by changing the times of feeding and running. Only four animals which learned the problem were given these trials. The order of correct responses was LRLR, i.e., the responses were motivated by 3, 22, 3, and 22 hours of hunger. In each of the four trials, whether the subjects chose correctly or incorrectly, a second trial was given to force the animals to respond in the direction opposite that chosen on the first trial.

B. *Results of Experiment II*

The data presented in Table II show that the albino rat can learn to

respond differentially to the same external situation using as cues for these responses various degrees of single drive. Each of the animals appearing in Table II ran six successive correct runs on six successive critical free trials (three at three hours of hunger and three at 22 hours of hunger). Including the intervening non-critical free trials, these animals had run to the correct side 12 successive times using the internal hunger drive as a cue.

Assuming equal likelihood of the animals' making right and left turns at the choice point, the probability that an animal will run in the prescribed pattern of six responses on critical trials alone on the basis of chance is about six in 100. (Actually, the likelihood of either a right or left turn at the choice point is less than .5, since in the critical trials reward was given in a previous training series on the side opposite that rewarded in the critical trial itself. On the early trials the animals were running between 90 percent and 100 percent of the time incorrectly on critical free trials. Thus we are being conservative in our estimate of significance.)

In the further critical trials, when the feeding and running times were changed, only one animal chose the correct goal box in all four trials. One animal responded correctly in three out of the four trials, and the remaining two subjects chose the correct goal box only once.

Discussion of Results

A. *General difficulties of learning the problem*

As mentioned previously in the introduction to this paper, it might have been expected that learning would occur if the different degrees of the single drive served as distinct stimuli, thereby resulting in discriminable stimulus complexes, one associated with the 3-hour motivation and the other associated with the 22-hour motivation. Despite the fact that Hull (1) used two distinct drives, hunger and thirst, in his experiment, learning was extremely difficult for his subjects. Leeper (4) facilitated learning in his study by permitting his animals to enter the inappropriate goal box and 'see' the contents therein. Here, also, two distinct drives operated. Kendler was also successful in obtaining discriminative behavior with the further complication of simultaneous hunger and thirst drives motivating his animals during training, one drive being rewarded on one side of the maze, and the second drive on the opposite side.

In the studies cited above, it should be apparent that differential responses were learned on the basis of distinct drive states serving as cues. Further, the goal objects in each case, were distinct, i.e., a right turning response became associated with one type of reward while the left turning response became associated with a different kind of reward. Those factors which facilitated learning in Kendler's study, and which were also operating in the Hull and Leeper experiments, were absent in the present experiment. Both sides (responses) became associated with a common reward and in the same amount. The fact that two of the animals never learned the problem in the second experiment and only one subject reached the criterion in the first experiment would lead to the conclusion that discrimination was extremely difficult. Many reinforcements were necessary even for those animals which did learn. Apparently different degrees of a single drive, even when they are as distinct as one would expect 3 and 22 hours of hunger to be,

are nevertheless difficult to discriminate. Hunger and thirst, on the other hand, while distinct not only in the regions of the body which they affect, are also distinct in the rewards which satisfy them. Thus, in a problem using both drives, learning should be *relatively* easy, in a problem using a single but varying drive with a common reward learning should be, and was, *relatively* difficult.

B. *Discussion of the differences in results found in Experiment I and Experiment II*

Learning was apparently facilitated by the changes introduced in Experiment II. Five animals of Exp. II had learned the problem in the 130 trials. Only one animal of Exp. I had learned after 130 trials in Exp. I. The reasons for this difference unfortunately cannot be independently identified because a number of changes were made between Exp. I and Exp. II. Any or all, or possibly none, of these changes facilitated learning. The changes made will be repeated in this discussion and the reasons for them will be mentioned.

1. Change in the amount of food given at the daily feeding

The seven gm. of food given daily in Experiment I was insufficient to satiate the animals even temporarily and they appeared to be constantly hungry. The subjects in the first experiment had usually finished eating in one-half hour, i.e., they had eaten all the ration in this period of time. With continuous hunger, it is possible that it was difficult to discriminate 3 from 22 hours of hunger, if, indeed, such distinct drives existed at all. In the second experiment, by giving the animals all the food they could eat for one hour, at least a temporary satiation was achieved. A good indication of this was the cessation of eating by the majority of the subjects in from 50 to 55 min.

2. The addition of a V Door to the maze

This door was installed in the maze in order to force the animals to hesitate at the choice point before entering either alley. Muenzinger and Gentry have shown that an increase in vicarious trial and error behavior at the choice point preceded mastery of the problem to be learned in their experiment (**6**). These authors, using rats, noted that the more often an animal stopped at the point of choice, the better was the learning. Muenzinger, in a more recent paper (**5**) in which he refers to his earlier study, states (5, p. 12):

> Whatever experimental conditions cause an animal to pause at the point of choice and indulge in VTE behavior will, thereby, accelerate learning.

3. A reduction in non-reinforced trials in each training series

As noted above, the problem to be learned was a difficult one. One factor operating to cause this difficulty was the fact that during a particular sequence the side of non-reward was being 'unlearned' or inhibited as a desirable response, e.g., a left turning response may become a response to be avoided on future runs—then on the following trial series the left turning response becomes the rewarded side and the right turning response was a 'to be avoided' response. On critical test trials the animal to run correctly must turn in opposition to his last rewarded response and in the direction in which he had previously found no food. In Experiment II the number of non-reinforced forced trials were reduced to minimize the inhibition built up for any particular side.

The same problem and the same reasoning follows when the factor of secondary reinforcement is considered, i.e., the non-rewarded side yielded some secondary reinforcement (having been previously associated with reward) on each forced run. This would mean that rather than having a simple learning problem of reinforcement versus non-reinforcement, the organism must, on any one trial, choose between primary reinforcement or secondary reinforcement. Reducing the non-reinforced trials reduced the secondary reinforcement.

Which of the factors was most effective in decreasing the problem difficulty or the degree cannot be answered. It can, however, be said that reducing the non-reinforcing trials would decrease the effect of both the counteracting effect of inhibition and secondary reinforcement.

C. *Discussion of results on critical trials with times of feeding and running changed*

As was noted in the results of Experiment II, with four rats the feeding time was modified (shifted from the regular feeding time of 9 and 10 A.M. to 12 P.M.). The animals were then run at intervals of 3 and 22 hours of hunger for four critical runs. The results approximated chance running.

A number of interpretations may be placed on these results. The results may be interpreted as indicating that the animals had not learned on the basis of internal hunger conditions but rather had learned on extra-environmental cues. A careful examination of the laboratory conditions by the experimenters, however, leads them to reject this explanation. So far as was discernible from as careful a survey as possible, the environmental conditions at the times of running were constant or any environmental changes were random in regard to running times.

Further interpretations lead to speculations concerning the nature of the stimulus complex differentiating 3 from 22 hrs. of hunger. Speculations concerning the 'time sense' in man in studies more directly concerned with this problem have led to few definitive conclusions. When it is recalled that the animals of this experiment had been fed each day at the same hour eight weeks prior to the change in feeding time, a conditioning of certain physiological rhythmic conditions to running times seems possible. If such were the case, the disruption of this patterning of the stimulus complex by the change in feeding time and a running of the animals the following day would lead to a prediction of the resultant chance behavior of the animals. A reconditioning of the feeding patterning would have shed some light on this matter, but unfortunately hindsight continues to be superior to foresight.

Summary and Conclusions

Two experiments were designed in an effort to discover to what extent the albino rate is capable of making discriminable responses on the basis of two strengths of a single drive serving as cues.

In the first experiment, 10 male and 2 female albino rats were trained in a T-maze for a period covering six weeks. They were rewarded for a response to the right when motivated by three hours of hunger, and for a response to the left when motivated by 22 hours of hunger. This series of correct responses was a prearranged irregular series to prevent ordinal patterning. Each response was rewarded 130 times, and an equal num-

ber of non-reinforced trials were run. Only one animal succeeded in learning the problem at the end of 90 reinforcements of each response. All other animals failed to learn the problem, and the first experiment was terminated at the end of 130 reinforcements. of each response (R & L).

In the second experiment, 12 female albino rats were trained in the same T-maze for a period covering eight weeks. Certain changes, which it was believed would facilitate learning, were introduced in the apparatus, and in the feeding and the training procedure. Ten of the 12 animals succeeded in reaching a criterion of learning which was shown to be significant.

Further critical trials with feeding and running times changed caused disruption of the habits established, but it is not believed that this in any way nullifies results obtained previously.

We may, therefore, conclude that the albino rat can learn to respond differentially to a given objective situation according to differential drive stimuli arising from various levels of intensity of a single drive. It was apparent, however, that the problem to be learned was a difficult one when compared with other learning conditions in the T-maze.

(Manuscript received
November 22, 1948)

References

1. Hull, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.
2. Hull, C. L. *Principles of behavior.* New York and London: D. Appleton-Century Company, 1943.
3. Kendler, H. H. Simultaneous hunger and thirst drives. Their influence upon the learning of two opposed spatial responses in the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
4. Leeper, R. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
5. Muenzinger, K. F. *Reward and punishment.* Univ. Colorado Studies, 1946, Vol. 27.
6. Muenzinger, K. F., & Gentry, Evelyn. Tone discrimination in white rats. *J. comp. Psychol.*, 1931, **12**, 195–205.

REINFORCEMENT AND EXTINCTION AS FACTORS IN SIZE ESTIMATION

BY WILLIAM W. LAMBERT, RICHARD L. SOLOMON, AND PETER D. WATSON

*Harvard University**

I. Introduction

In recent experiments on the psychology of perceiving, there has been a noticeable tendency to emphasize determinants which might be classed as motivational in character. The work of Sanford (**6, 7**) involving the relationship between drive states and 'autistic perceiving,' and the extension of this work by Murphy and his collaborators (**3, 5, 8**), and by McClelland and Atkinson (**4**), illustrate this trend. More closely related to the present problem is the work of Bruner, Postman, and their collaborators (**1, 2**) dealing with the 'selection' and 'accentuation' of perceived objects relative to the 'value systems' of an individual. Two of their experiments in particular illustrate the operation of the conceptualized value dimension. Bruner and Postman (**2**) found that circles of the same diameter, embossed with (1) a high-valued social symbol, and (2) a low-valued social symbol, were judged to be larger than circles embossed with (3) a neutral symbol. This might indicate that 'perceptual accentuation' is a U-shaped function of a value dimension varying from -1 to $+1$, with a minimum of accentuation at 'neutrality.' Bruner and Goodman (**1**) have shown that poor children tend to overestimate the size of coins more than rich children do. These experimenters stated: "The reasonable assumption was made that poor children have a greater subjective need for money than rich ones." (**1**, p. 39) They further asserted that "the greater the value of the coin, the greater is the deviation of *apparent* size from *actual* size." (**1**, p. 38)

The multitude of influences correlated with being rich or poor makes it difficult to analyze the specific determinants of size overestimation. It was thought that some light could be shed on this problem by experimentally controlling the life history of children with respect to an initially neutral object. Specifically, we wished to associate a relatively neutral poker chip[1] with candy reward and later extinguish this association by removal of reward and to measure the effects of such procedures on the estimated size of the poker chip. Our hypothesis was that 'value,' as defined by changes in apparent size, is a function of both reinforcement and extinction procedures.

II. Subjects and Procedure

In the first study, 32 children from the Harvard Nursery School (ages three to five) were divided into 22 experimental subjects and 10 control subjects. In the second study, 22 children of comparable age from a Salvation Army Nursery School provided 15 experimental subjects and 7 control subjects.

The experimental subjects were individually introduced to a token-reward situation where they turned a crank 18 turns in order to receive a white poker chip which, when put into a slot,

* This research was facilitated by the Laboratory of Social Relations, Harvard University. The authors wish to thank Miss Winifred Lydon, Director of the Harvard Veteran's Nursery School, and Major Gertrude Atkinson, of the Salvation Army Nursery School, Boston, for their indispensable help and cooperation in carrying out this study.

[1] Only one of our children knew what a poker chip was. It was called a circle in our experiment.

led to the automatic delivery of a piece of candy. The control subjects were introduced into the same situation, but candy came directly after work, *without* the mediation of a poker chip. In the first study, both groups worked (and were rewarded) once a day for 10 days; in the second study, the subjects worked (and were rewarded) *five* times a day for 10 days.

Size estimates of the white poker chip token were made by the subjects (1) prior to the experiment; (2) after 10 days of reward; (3) after extinction had occurred (11th day); and (4) after reward had been reinstated (12th day).

Measurements were taken with the equipment designed and used by Bruner and Goodman (1). This equipment was composed of a rectangular wooden box (9 × 9 × 18 in.) with a 5-in. square ground-glass screen in the center of the front panel, and a control knob at the lower right-hand corner. At the center of the ground-glass screen the subject was presented with a circular patch of light (16.2 app. ft. cdls.) the diameter of which was under the control of the knob. The light source was a 60-watt incandescent light shining through an iris diaphragm which could be varied (in terms of the visible light patch) from ¼ to 2 in. As Bruner and Goodman reported: "The circle was not truly round, containing the familiar nine-elliptoid sides found in the Bausch and Lomb iris diaphragm. It was so close, however, that subjects had no difficulty making the subjective equations required of them." (1, p. 37)

The subjects stood in front of the apparatus with the light patch at or slightly below eye level, and about 12 to 18 in. away. The token, pasted on a 5-in. square gray cardboard, was held by the experimenters so that it was parallel to the circular patch. About 7 in. separated the centers of the two objects to be compared.

The judgment problem was presented to the children of both groups as a game. Each child made his estimates alone. Two judgments starting from the open and two starting from the closed position of the iris were obtained from each child at each measurement session; these judgments were made in an order which was counter-balanced for direction of turning of the control knob. The children were not informed of their success in approximating the actual size of the poker chip.

On the 11th day—after 10 days of rewarded trials—extinction was instituted. The children of both groups worked, but no candy was forthcoming. They worked until they met the arbitrary criterion of extinction: three min. during which they did not turn the handle of the work machine. The size estimates were made immediately after the subject had met the extinction criterion.

On the 12th day the subjects were reintroduced to the reward sequence, and the work brought candy to the control group and token plus candy to the experimental group. Size estimates were made immediately after this 12th session.

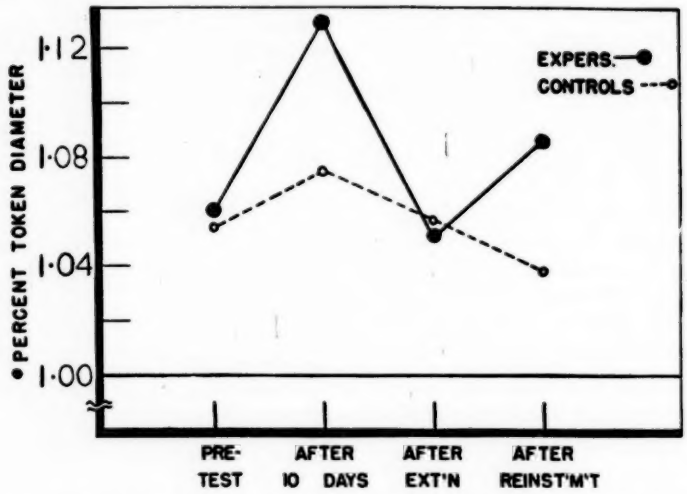


FIG. 1. Effects of the experimental conditions upon children's estimates of the diameter of a token when these estimates are taken as percents of the true diameter

III. Results

The results for both nursery schools were combined and they are shown graphically in Fig. 1. The four size estimation sessions are distributed on the x-axis; the mean estimate of the token size in terms of percent of actual size is shown on the y-axis. The actual size is indicated by the horizontal line parallel to the x-axis. The actual size is indicated by the horizontal line parallel to the x-axis. The means for the experimental group are connected by the solid lines, and the means for the control group are connected by the dotted lines. The connecting lines are meant to increase legibility; they do not imply a continuous function of any sort.

It would appear that the control group showed no significant changes with experience. The experimental group, however, showed a rise in the apparent size of the token after ten days of using the token to obtain reward. The estimates dropped to the level of the beginning estimates following the extinction procedure in which the token no longer led to candy reward. The estimates went back in the direction of over-estimation when reward was reinstated on the 12th day.

The mean size estimates in arbitrary units of the comparison-stimulus diameter are given in Table I, together with the corresponding percent of the actual token diameter, for each of the four points in our experiment. The results for our two studies are combined, since there were no appreciable differences between the 10-reinforcement and the 50-reinforcement experiments.

Analyses of variance[2] were performed on the data which are summarized in Table I. The following differences are of interest: (1) In the experimental group, the estimated size of the token after 10 days of reinforcement was significantly greater than at the pretest. This difference is reliable at the one percent level of confidence. (2) In the experimental group, the size estimates after extinction were significantly smaller than they were after the 10 days of reinforcement. This difference is reliable at the one percent level of confidence. (3) In the experimental group the rise in estimated size following reinstatement of reward is significant at the one percent level of confidence. (4) In the control group, none of the four mean estimates is significantly different from any other. (5) The mean estimates for the experimental and control groups after ten days of reinforcement are significantly different from one another with a reliability between the one and five percent levels

TABLE I

The Alteration of Size Estimation with Experience in the Experimental Situation

Experimental Group

	Mean Estimated Size*	σ_m	Percent Actual Size
Pretest	66.8	1.2	1.06
After 10 days	70.9	1.1	1.13
After extinction	66.3	1.3	1.05
After reinstatement	68.5	1.8	1.09

Control Group

	Mean Estimated Size*	σ_m	Percent Actual Size
Pretest	66.4	1.5	1.05
After 10 days	67.7	1.2	1.07
After extinction	66.6	1.2	1.06
After reinstatement	65.4	1.4	1.04

* Actual size of piker chip is 63.0 in arbitrary units of diameter. The error of measurement of diameter by experimenter is ±0.2 units.

[2] These analyses are uncorrected for the correlation between successive sets of estimates. They are thus conservative.

of confidence. (6) The mean estimates for the experimental and control groups after reinstatement of reward are not significantly different from one another even though a marked trend seems evident.

IV. Discussion

Several alternative theoretical interpretations for our results could be made. Since experiments are in progress to study further the factors involved, these possibilities will merely be listed at this time. These views are not mutually exclusive, but overlap, as do so many formulations in this field.

1. The estimation changes in the experimental group may be compatible with a generalized pattern of behavior which we could call the 'cookie effect.' That is, the effect may be peculiar to our culture where, for example, a 'bigger cookie is *better* than a little one.' 'Bigness' and 'value,' or 'bigness' and 'goodness,' may be acquired equivalencies for our children, particularly at the ages of the subjects used here. Experiments have been planned which may provide evidence on whether this phenomenon is 'culture bound' or not.

2. These results may provide a measure reflecting some of the secondary reinforcing characteristics taken on by the token during the reinforced trials. These characteristics become lost when reinforcement is not maintained, as during extinction, but are restored when reward is reinstated. This formulation, if further bulwarked with evidence, could serve to integrate perceptual distortion phenomena with learning theory and possibly provide a valuable indirect measure of secondary reinforcement.

3. It is possible that the size enhancement phenomenon can provide us with inferences about perceptual processes as envisioned by Bruner and his collaborators (1, 2). They hypothesize: "The greater the social value of an object, the more will it be susceptible to organization by behavioral determinants." (1, p. 36) In its learning aspects, however, overestimation of size may reflect either 'expectancy' or 'hypothesis' formation (and decay) or it may, as stated above, reflect learned 'needs' which operate in the workings of this conceptualized perceptual process. The actual mechanism which produces overestimation following reinforcement is, however, entirely obscure at the present stage of our research.

In view of the fact that relatively 'neutral' poker chips were used in the experiment, our data cannot be legitimately compared with the coin size data of Bruner and Goodman (1). In addition, our two nursery school groups do not fulfill the criteria of distinct economic class differences. In no sense can we call one group 'rich children' and the other group 'poor children.'

It is interesting to note the possibility that effects such as those discussed here depend on a 'difficult' or 'ambiguous' judgment situation. Probably, the more ambiguous the stimulus situation, the more strongly can reinforcement and motivational factors operate in determining size judgments.

4. It is interesting to note that, following extinction procedures, the estimates of the experimental group do not increase above the original level, when the chip was 'neutral.' This could mean that the U-shaped function postulated to relate accentuation and value does not apply here. Or it could mean that extinction removes positive value without producing negative value. Perhaps extinction by punishment is necessary for produc-

ing negativity and an increase in size estimates at the negative end of the U-shaped function.

V. Summary

We have described the results of an experiment which was designed to investigate the effects of reinforcement and extinction on size estimation. It was found that the establishment of a token reward sequence results in relative overestimation of the token size. Extinction of the sequence removes this overestimation tendency to a great extent. The results are thought to have relevance for both learning and perception theory.

(Manuscript received
November 29, 1948)

References

1. Bruner, J. S., & Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
2. Bruner, J. S., & Postman, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
3. Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
4. McClelland, D. C., & Atkinson, J. W. The projective expression of needs: I. The effect of different intensities of hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.
5. Proshansky, H., & Murphy, G. The effects of reward and punishment on perception. *J. Psychol.*, 1942, **13**, 295–305.
6. Sanford, R. N. The effect of abstinence from food upon imaginal processes; a preliminary experiment. *J. Psychol.*, 1936, **2**, 129–236.
7. Sanford, R. N. The effect of abstinence from food upon imaginal processes; a further experiment. *J. Psychol.*, 1937, **3**, 145–159.
8. Shafer, R., & Murphy, G. The role of autism in a visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335–343.

SEMANTIC AND PHONETOGRAPHIC GENERALIZATIONS OF SALIVARY CONDITIONING TO VERBAL STIMULI

BY GREGORY RAZRAN

Queens College

In a previous preliminary study by the writer (**32**), it was found that in adult human *Ss* salivary CR's to words generalized to both synonyms and homophones, the generalization to the homophones being about 37 and to the synonyms about 59 percent. This finding was practically duplicated by another experimenter using the same stimulus words (*style*—'fashion,' 'stile'; *urn*—'vase,' 'earn'; *freeze*—'chill,' 'frieze'; *surf*—'wave,' 'serf') but another response, the GSR (**35**); as well as in the main corroborated by a few other studies that are in some ways related to the writer's investigation (**36, 45; 4, 6, 16**).

In the present study[1] an attempt was made to map out the entire range of CR generalizations to a variety of words that were in some ascertainable way related to the conditioned words, either along the dimension of 'meaning-content' or along that of 'sound-spelling.' It was thought that generalizations along the 'meaning-content' dimension could best be studied by testing CR carry-overs from conditioned words to their supraordinates, subordinates, coordinates, contrasts, predicatives, whole-part's, part-whole's, and compound words. Such generalizations might be termed semantic, although the term dimension is perhaps begging the question here. On the other hand, it was thought that generalizations along the dimension of 'sound-spelling' could be investigated by determining CR transfers to words having varying numbers of sounds—more correctly, phonemes—and letters that are identical with those in conditioned words. Such generalization has been named phonetographic, since it is obviously due to a relatedness in both auditory-phonetic and visual-graphic elements[2]—a relatedness that is furthermore usually unequal in English words. In addition, to complete the series, it was considered worth while including some generalization words that were related both semantically and phonetographically to the conditioned words, such as 'signal' and 'signature' in relation to *sign*, as well as pairs of words with similar photetographic yet different semantic relatednesses, such as *bake*—'book' vs. *take*—'took,' and *mock*—'match' vs. *lock*—'latch.' In all, the program was rather extensive, and about some of its aspects it was hoped to gain only fragmentary information.

Procedure

The writer's technique of salivary conditioning has been described previously (**30–32**), and need not be repeated in all details. Essentially, it consists of measuring saliva by weighing increments in dental cotton rolls (Johnson and Johnson, No. 3, 0.5 × 1.5 in.) inserted under the *Ss*' tongues for short periods of time, usually for one min. Since the cotton-in-the-mouth is by no means an inactive stimulus, periods of control salivation must be rotated with periods of experimental salivation, and, of course, pretraining salivation to the stimuli-to-be-conditioned must be determined. To prevent evaporation, scale corrosion, and absorption, the rolls are first weighed in small cellophane envelopes and then re-weighed in the envelopes

[1] This study is being prepared for publication while the writer is a Fellow of the John Simon Guggenheim Memorial Foundation.

[2] Failure to take in consideration graphic relatedness is illustrated in one study (**6**) in which the homophones of 'vein'—'vain' were used alongside those of 'knows'—'nose.'

immediately after their removal from the *Ss'* mouths. A special scale, permitting instantaneous weighing accurately to two mg., was constructed for this purpose by Roller Smith and Company, Bethlehem, Pa.

The stimuli-to-be-conditioned are presented a number of times, intermittently, for periods of one to two sec., while the *Ss* are eating small 8-shaped pretzels or tea-sandwiches, or sucking at lollipops or peppermint candy (chewing gum has proved less effectual). The eating periods last from two to four min., with one- to two-min. rest intervals between them, and 10–15 of such eating periods constitute an experimental training session. The training sessions are scheduled in late mornings or afternoons, and the food is varied from one eating period to another in the same sessions. The multiple intermittent presentations of stimuli-to-be-conditioned are intended to secure a maximum of attention for the stimuli; the lengths of the eating periods purport to make the entire task more 'molar' and meaningful; and the varying of the food and the timing of the sessions are calculated to insure adequate psychophysiological motivation.

A very special characteristic of the writer's technique is that of misinforming the *Ss* about the nature of the experiment. This misinforming is a *sine qua non*, if disrupting subjective attitudes in the wake of the *Ss'* knowledge-expectancy of being conditioned are to be controlled. Fortunately, with the method outlined, disguise has not been too difficult. Telling the *Ss* that the experiment aims to study "the effects of hunger upon eye-fatigue" or 'upon memory,' and actually giving a few 'sham' memory or eye-fatigue tests, has been most commonly and successfully used in the last few years. Occasionally, an *S* will 'catch on,' and then the practice has been to assign him to another study and to pledge him not to tell his colleagues. 'Catching on' is of course a rarity with *Ss* unfamilar with academic psychology.

In the particular procedure of the present experiment, a number of discrete English words —five in one part of the experiment, seven in another—were flashed one at a time and intermittently on a screen, in random order, for periods of one sec., during eating periods of three min. The intervals between eating periods varied from one-half to one and one-half min. and those between flashes ranged from one-half to one and one-half sec. Fifteen eating periods constituted a training session, and each training session was followed by a training-testing session. In the training-testing sessions, five 3-min. eating periods alternated with an equal number of 21-min. testing periods; and during each testing period the amount of conditioned salivation to two conditioned words, four non-conditioned but related generalization words, and one non-related and non-conditioned control word, was determined by the cotton technique in mg. The eight *Ss* in the experiment—all college undergraduates—were divided into two equal Groups, I and II, which, in turn, were sub-

TABLE I

PLAN OF THE EXPERIMENT. CONDITIONED AND GENERALIZATION WORDS USED IN EACH TRAINING AND IN EACH TRAINING–TESTING SESSION FOR EACH OF THE EIGHT *Ss*. Conditioned words are italicized.

Experimental Group	Experimental Sessions				
	1	2	3	4	5
Ia	*Dog, Dark, Flower, Sign, Yankee*	*Dog, Flower:* Animal, Bark, Glower, Pansy	*Dark, Yankee:* Afraid, American, Doodle, Light	*Dog, Sign:* Cat, Signal Signature, Terrier	*Dark, Flower:* Mark, Room, Shower, Petal
Ib	As above, in Ia	*Dark, Flower:* Mark, Room, Shower, Petal	*Dog, Sign:* Cat, Signal, Signature, Terrier	*Dark, Yankee:* Afraid, American, Light, Doodle	*Dog, Flower:* Animal, Bark, Glower, Pansy
IIa	*Bake, Day, Goat, Lock, Mineral, Mock*	*Day, Goat:* Horns, May, Night, Week	*Bake, Take:* Book, Give, Sun, Took	*Lock, Mock:* Dock, Key, Latch, Match	*Goat, Mineral:* Animal, Copper, Sheep, Water
IIb	As above, in IIa	*Goat, Mineral:* Animal, Copper, Sheep, Water	*Lock, Mock:* Dock, Key, Latch, Match	*Bake, Take:* Book, Give, Sun, Took	*Day, Goat:* Horns, May, Night, Week

divided into two equal Subgroups, I*a* and I*b*, and II*a* and II*b*. Group I experimented with a set of five conditioned words and 16 generalization words, and Group II with another set of seven conditioned words and 16 generalization words, while the Subgroups in each Group differed from each other only in the sequences of the testings of their CR's. There were 20 CR determinations for each generalization word, five from each of four *S*s, but some conditioned words were tested 40 times. The entire scheme of the experiment is presented in Table I.

The data in Table I may well be summarized, for the sake of clarity, in the text. They show that the conditioned words in Groups I*a* and I*b* were: (1) *Dog*, with the generalization words of 'animal,' 'bark,' 'cat,' and 'terrier'; (2) *Dark*, with the generalization words of 'mark,' 'light,' 'afraid,' and 'room'; (3) *Flower*, with the words of 'glower,' 'shower,' 'petal,' and 'pansy'; (4) *Sign*, with 'signal' and 'signature'; and (5) *Yankee*, with the generalization words of 'American' and 'Doodle.' In Groups II*a* and II*b*, the conditioned words were: (1) *Day*, with the generalization words of 'may,' 'night,' and 'week'; (2) *Goat*, with the words of 'animal,' 'horns,' and 'sheep'; (3) *Mineral*, with 'copper,' and 'water'; (4) *Bake*, with 'book' and 'sun'; (5) *Take*, with 'took,' and 'give'; (6) *Lock*, with 'key,' and 'latch'; and (7) *Mock*, with the words of 'dock,' and 'match.'

The relatednesses of the generalization words to the conditioned words will readily be recognized as those of the familiar 'association' categories: word derivatives, supraordinates, subordinates, coordinates, contrasts, part-whole's, whole-part's, agent-action's, action-agent's, attribute-substance's, cause-effect's, word compounding's. The relatednesses of 'match' to *mock* and of 'book' to *bake* has been termed as those of pseudo-derivatives, while the phonetographic relatednesses of 'glower' and 'shower' to *flower*, of 'dock' to *mock* and of 'mark' to *dark*, and of 'may' to *day* might perhaps be best specified as those of "3 of 4 common phonemes and 5 of 6 common letters," "2 of 4 common phonemes and 4 of 6 common letters," "2 of 3 common phonemes and 3 of 4 common letters," and "one of 2 common phonemes and 2 of 3 common letters," respectively. It might also be added here that words for food and eating, as well as phonetographically related words in semantic categories, were excluded from the sets.

Results

The results are presented in Tables II, III, and IV. They are given in percents of generalization, that is, the conditioned salivation to generalization words as percents of the condi-

TABLE II

Generalization of Salivary Verbal Conditioning to Semantically and Phonetographically Related Words.

Each entry is a mean of 20 one-min. measurements obtained from four *S*s (Group I).

Conditioned Words	Conditioned Salivation in Mg.	Generalization Word	Percent of Specific Generalization	Relatedness of Generalization Words to Conditioned Words
Dog	417± 8.5	Animal	22.2	Supraordinate
Dog	417± 8.5	Bark	56.8	Agent-Action (predicative)
Dog	417± 8.5	Cat	38.1	Coordinate
Dog	417± 8.5	Terrier	50.3	Subordinate
Dark	394± 9.7	Afraid	32.4	Cause-Effect (predicative)
Dark	394± 9.7	Light	40.1	Contrast
Dark	394± 9.7	Mark	36.1	2 of 3 common phonemes, 3 of 4 common letters
Dark	394± 9.7	Room	13.2	Attribute-Substance (predicative)
Flower	432±10.7	Glower	35.1	3 of 4 common phonemes, 5 of 6 common letters
Flower	432±10.7	Pansy	42.2	Subordinate
Flower	432±10.7	Petal	45.1	Whole-Part
Flower	432±10.7	Shower	20.2	2 of 4 common phonemes, 4 of 6 common letters
Sign	424± 9.8	Signal	69.2	Word Derivative
Sign	424± 9.8	Signature	65.1	Word Derivative
Yankee	387± 9.6	American	24.1	Supraordinate
Yankee	387± 9.6	Doodle	13.1	Compound Word

TABLE III

GENERALIZATION OF SALIVARY VERBAL CONDITIONING TO SEMANTICALLY AND PHONETOGRAPHICALLY RELATED WORDS.

Each entry is a mean of 20 one-min. measurements of four *S*s (Group II).

Conditioned Words	Conditioned Salivation in Mg.	Generalization Word	Percent of Specific Generalization	Relatedness of Generalization Words to Conditioned Words
Bake	487± 9.9	Book	18.1	Pseudo-Derivative
Bake	487± 9.9	Sun	31.8	Action-Agent (predicative)
Day	453± 7.8	May	19.6	One of 2 common phonemes, 2 of 3 common letters
Day	435± 7.8	Night	41.2	Contrast
Day	435± 7.8	Week	39.8	Part-Whole
Goat	399±14.7	Animal	22.3	Supraordinate
Goat	399±14.7	Horns	32.7	Whole-Part
Goat	399±14.7	Sheep	37.2	Coordinate
Lock	418±10.7	Key	39.2	Coordinate
Lock	418±10.7	Latch	51.2	Word Derivative
Mineral	409±11.2	Copper	35.4	Subordinate
Mineral	409±11.2	Water	12.2	Compound Word
Mock	376± 9.2	Dock	27.2	2 of 3 common phonemes, 3 of 4 common letters
Mock	376± 9.2	Match	8.1	Pseudo-Derivative
Take	398± 9.7	Give	40.2	Contrast
Take	398± 9.7	Took	72.4	Word Derivative

tioned salivation to conditioned words. These percents are furthermore termed percents of *specific* generalization, inasmuch as they represent percents of conditioned salivation to related generalization words over and above those to non-related control words. Table II details these percents for each generalization word in the experiment with Group I; Table III presents them similarly for the experiment with Group II; and Table IV combines the data from both Groups, and also groups the data for each generalization word into word categories, of the traditional 'association' variety.[3] The number of words in each category is rather small, in no case more than four; but one should consider the fact that each word was tested 20 times. The *t*'s and *P*'s for the differences in generalization were ascertained for all entries in the tables but their small range did not seem to warrant the expense of separate tabulations in extra tables. Roughly, differences of six or seven percent among the entries in Tables II and III and of five or six percent among those in Table IV could well be considered fully significant statistically, with P equalling 0.01.

In order of definitiveness and expectations, the findings in Tables II, III, and IV seem to arrange themselves somewhat as follows. First, there is the large variation in amount of generalization to words of varying relatednesses (and the apparent fact that the salivary CR is able to detect

[3] These word categories are, to be sure, as yet in a stage of crude classification. Consider the words 'petal' and 'horns' in a 'whole-part' category, and the words 'dog' and 'man' in an 'agent-action' category. Clearly, 'petal' is related to fewer 'wholes' and 'dog' to fewer 'actions' than are respectively 'horns' and 'man,' so that neither CR generalization nor association speed and frequency could be expected to be equal in the two cases (they were not equal in the present experiment).

TABLE IV

PERCENT OF SPECIFIC GENERALIZATION OF SALIVARY VERBAL CONDITIONING TO SEMANTICALLY AND PHONETOGRAPHICALLY RELATED CATEGORIES OF WORDS. Each entry in the third column is a mean of 20–80 measurements (20 times the entry in the second column).

Word Category	No. of Words in Category	Percent of Specific Generalization
Word Derivative	4	64.5
Subordinate	3	42.6
Contrasts	3	40.5
Part-Whole	1	39.8
Whole-Part	2	38.8
Coordinate	3	38.2
3 of 4 common phonemes, 5 of 6 common letters	1	35.1
Predicative	4	35.0
2 of 3 common phonemes, 3 or 4 common letters	2	31.6
Supraordinate	3	22.9
2 of 4 common phonemes, 4 of 6 common letters	2	20.2
One of 2 common phonemes, 2 of 3 common letters	1	19.6
Pseudo-Derivative	2	13.1
Compound Words	2	12.6

such relatednesses). Thus, the generalization form *mock* to 'match' was 8.1 percent and that from *take* to 'took' was 72.4 percent. Second, there is the crude gradient of generalization to phonetographically related words. The generalization CR from *flower* to 'glower' was 35.1 percent, from *dark* to 'mark' 33.2 percent, from *mock* to 'dock' 29.6 percent, from *flower* to 'shower' 20.2 percent, and from *day* to 'may' it was 19.6 percent. Third, there is the nearly equal CR generalization to 'contrasts,' 'coordinates,' 'subordinates,' 'part-whole's,' and 'whole-part's.' The near-equality of the generalization to the first two categories—contrasts and coordinates—could well be expected, since the frequencies of these categories in free association and their reaction times in controlled association tests have, likewise, been nearly equal (**43, 44; 15, 17, 23, 25, 42**). But, on the other hand, the high degree of CR generalization to subordinates, part-whole's, and whole-part's differs considerably from results of association studies which had shown these categories to have lower frequencies and longer reaction times than contrasts and coordinates (*op. cit.*). And even more at variance with association studies and with psychological principles of habits, in general, is the finding that CR generalizations to supraordinates were only about one-half of those to subordinates. Finally, the most unexpected results seem to be those that show so little generalization to the second word of a compound word, such as from *mineral* to 'water' and from *Yankee* to 'Doodle.'

The last two findings have led the writer to a considerable amount of theorizing. And in his theorizing, he formulated a new hypothesis—really not altogether new, since he had an inkling of it since 1930 (**29**)—of generalization, in general—and designed an experiment to test the hypothesis. Both hypothesis and confirming experiment will be taken up in the next two sections, under 'Discussion, Theory, and a Supplementary Experiment' and under 'Generalization, in General.'

DISCUSSION, THEORY, AND A SUPPLEMENTARY EXPERIMENT

In 1930, the writer stated the following with regard to CR generalization: "Generalization of CR's is well accounted for by irradiation, which apparently acts in two ways. First, there is considerable irradiation

from the focus of the active conditioned stimulus, *S*, to the foci of similar inactive stimuli, s_1, s_2, s_3 . . . s_n, so that there are established not only the main pattern *s*-O-R, but also patterns s_1-O-R, s_2-O-R, s_3-O-R . . . s_n-O-R, the strengths of which fall off as their anatomical and functional contiguity with the active conditioned stimulus become remote. Secondly, there is irradiation from the foci of the similar stimuli to the active stimulus so that the main pattern *s*-O-R is also activated in some degree by the application of the similar stimuli" (29, p. 38).

Leaving out the then-more-fashionable pseudo-physiological phrasings such as 'anatomical and functional contiguity' and the youthful optimism of "is well accounted for," and using 'irradiation' behaviorally as synonymous with 'spread,' the quoted paragraph may now be reformulated as follows. There may be two basic hypotheses about the mechanism of CR generalization: one, that this generalization is a function of the *original CR training situation* and is directionally determined by the relatedness of the generalization stimuli to the conditioned stimuli; two, that this generalization is a function of the *subsequent CR testing situation* and is directionally determined by the relatedness of the conditioned stimuli to the generalization stimuli. According to the first hypothesis, if an *S* has formed a CR, let us say, to a tone of 700 cycles, there are formed thereby automatically, through 'spread' or 'irradiation,' also CR's to generalization stimuli 600, 500, and so on, 800, 900, and so on. These generalization CR's are there, in the organism, latent but ready to function; there is a true spread of learning, a true associative or CR gradient. According to the second hypothesis, the *S* has formed a CR to only a particular stimulus, to only a tone of 700 cycles; but upon subsequent testing with generalization stimuli, with other tones, there takes place a spread or irradiation from the generalization stimuli to the conditioned stimulus, thus activating the conditioned response. There is no true spread of learning, no latent bonds, and no true associative or CR gradient—but only a stimulus gradient and a stimulus or sensory generalization, using the last two terms, not as a CR stimulus or sensory generalization, but in their true traditional sense exemplified by nineteenth century psychophysical and psychophysiological studies.

Now, the first hypothesis has been made the basis of the theories of Hull and Spence in this country (11–14, 38–40) and of the views of Pavlov, Bekhterev and their many students in the U.S.S.R. (1, 2, 26–28, and a multitude of others too numerous to cite here.) But it is the second hypothesis that is favored by the law of parsimony and that fits best the data of the present study. As will be remembered, the chief discordance of the present study was the finding that there was twice as much CR generalization to subordinate as to supraordinate categories, and that there was hardly any generalization from the first to the second word in a compound of two words. But the categories named subordinate in accordance with the first hypothesis of 'relatedness of generalization stimuli to conditioned stimuli' are really supraordinate in accordance with the second hypothesis of 'relatedness of conditioned stimuli to generalization stimuli,' and the supraordinates are by the same reasoning subordinates. Likewise, presenting the second word of a compound of two words as a generalization stimulus means, ac-

cording to the first hypothesis, bringing into play a natural speech habit, but, according to the second hypothesis, it means invoking an unnatural inverted speech habit. To specify: testing for CR generalization to the words 'terrier' and 'animal' in *S*s conditioned to the word *dog* means testing for subordinate and supraordinate categories, respectively, in accordance with the first hypothesis. But, according to the second hypothesis, it means testing for supraordinates and subordinates—and it is, of course, the CR generalization to supraordinate categories that would be expected to be higher—and thus actually came out higher in the present study—than that to subordinate categories, in accordance with known facts of habit and association. Again, testing for CR generalization to the words 'water' and 'Doodle' in *S*s conditioned to the words *mineral* and *Yankee* means, according to the second hypothesis, invoking an inverted speech habit and backward associations of 'Doodle Yankee' and 'water mineral,' and it is no wonder that very little CR generalization occurred in this case. To test this last consideration, a special supplementary experiment was made at the end of the study.

In the special supplementary experiment, two *S*s were used, while the stimuli-to-be-conditioned were the words *doodle* and *water*. The conditioning procedure was the exact duplicate of the main experiment except of course that, instead of flashing five or seven words, only two words were flashed on the screen during eating periods. The salivary CR's became established, as usual, in one training session, and then the *S*s were tested for generalization to the words 'Yankee' and 'mineral' on three successive training-testing sessions. And, to be sure, the CR generalization was 57 percent for 'Yankee Doodle' and 49 percent for 'mineral water,' thus confirming the contention of the 'subsequent testing situation' hypothesis—the second hypothesis—of CR generalization that the direction of CR generalization is from the generalization stimuli to the conditioned stimuli, and not vice versa.[4]

Generalization, in General

The results of the present study have been obtained with a response the conditioning of which has been shown by the writer to be quite subject to central or verbal or conscious controls (30). And, while salivation is not as voluntary as winking or finger withdrawal in the sense that an *S* cannot just 'will' a gland to secrete saliva, but must invoke a word or thought to do it, and while the writer has attempted—and thinks he succeeded—to control central factors, it is nevertheless true that the present results should not be wholly generalized to those that, let us say, might be obtained with pupillary conditioning, or to those obtained from animal studies (cf. Lashley & Wade, 20). Moreover, the present study is only one of a great many CR generalization studies. The writer is thus tempted to express an *opinion* about generalization, in general, particularly after

[4] The rather high CR generalization to the three words in the whole-part and part-whole categories could probably be accounted for by the particular words used. In the whole-part category—using now categories in accordance with the second hypothesis—the conditioned word was *day* and the generalization word 'week,' a word that is related to *day* also as a common compound 'week day.' In the part-whole category, the generalization words were 'petal' and 'horns' and the conditioned words were *flower* and *goat*. The special semantic and associative closeness of 'petal' to *flower* needs no further discussion, while 'horns' produced a CR generalization of only 32.5 percent.

the recent exchange of views between Professors Lashley and Hull (**20, 13**).

In the writer's opinion, the contention that a true CR generalization gradient—so-called primary stimulus generalization—is developed during the original training of the conditioning, and is thereafter prepotent over, or even determinant of, subsequent experiential or secondary or mediate factors, cannot be considered as something that has been definitely established either in the Russian laboratories or in any laboratory in this country. There are, to be sure, studies the results of which might be construed as bona fide evidence for such a doctrine. But the number of studies that clearly do not fit the doctrine is considerably greater. Furthermore, *it is very much easier to interpret graded CR's in terms other than primary stimulus generalization* (Loucks, **22**; Lashley, **20**; Guthrie, **7**; Hilgard, **8**, p. 338; Konorski, **18**; the writer) *than it is to account for failures of CR gradients and still adhere to the postulate.* Such accounts are particularly difficult for results from the extremely routinized Russian laboratories, since failures there could hardly be attributed to inadequacies and variations of methodology, technique, or apparatus.

To specify the argument: The Russian laboratories have yielded to date, in the writer's estimate, about 100 studies containing data on CR generalization. Loucks very ably analyzed six of the studies (**21**). The writer has read most of the hundred, but, unfortunately, has not analyzed them, the way Loucks did, for irradiation, or even for gradients. However, he does have strong impressions. And, if these impressions have any validity, they are that *the Russian evidence* (the writer himself *tended* to accept the evidence at one time [**29, 33, 34**]) *for CR gradients of generalization could not meet the test of a statistical analysis.* The Russian results, the writer believes, would show that related stimuli *mostly* produce CR's without conditioning and that these CR's are *usually* weaker—smaller in magnitude, longer in latency, more rapid in extinction—than the CR's to the original conditioned stimuli. But no stable regular and lawful gradients of CR generalization to suit the views of Pavlov (**26–28**) and the postulates of Hull (**11–14**) appear. Some of the Russian results are indeed strikingly negative. The writer can think, for instance, of an experiment by Shevalev (**37**) with tactile CR's to faradic shock, one of the best controlled Russian experiments. One of Shevalev's dogs, in whom an apparently stable tactile CR gradient on the left hip had been developed, suddenly telescoped the entire CR area, including the exact spot of the application of the conditioned stimulus, from the hip to the tail region. Shevalev explains his results in Russian fashion by saying that the particular dog developed a special 'internal inhibition.' But such an account is of course hardly satisfactory, if not meaningless, especially when in the entire 204 pages of the study and in its 40 odd protocols nothing is mentioned of the stimulus for, or even the condition of, the 'internal inhibition.' [5]

If Russian results on CR gradients are spotty, the data from American laboratories are not only spotty but also very scanty. Space permits an analysis of only one American study, the one most often quoted, the study by Hovland with tone frequencies (**10**). It is a nice study. But besides the fact that Hovland worked with the GSR in adult human *S*s,[6] the evidence itself is very

[5] The present article was prepared a few months before another article by the writer (*Psychol. Bull.*, September, 1949) in which the total evidence of CR generalization from American and Russian laboratories has been subjected to a quantitative and critical analysis. The reader is referred to that article for more general findings and interpretations. The latter, however, do not in the main contradict but expand and supplement the more tentative and limited views expressed in the present article.

[6] The statement by Professor Hull about the 'relatively non-voluntary character of galvanic skin reactions' (**13**) does not really bear much upon the problem of stimulus generalization, since here it is the consciousness of the stimuli, and not only of the responses, that counts; and

far from conclusive. The mean magnitude of the CR to the conditioned stimulus in Hovland's 20 *S*s was 18.3± 0.57 (PE), while the mean magnitudes to the generalization stimuli—25, 50, and 75 j.n.d.'s removed from the conditioned stimulus—were 14.9 ± 0.64, 13.6 ± 0.8, and 12.9 ± 0.49, respectively (*op. cit.*, Table I). The differences are fully reliable statistically only between the conditioned stimulus and the generalization stimuli and also between the first and third generalization stimuli (Diff./$PE_{diff.}$ = 4.21). But the differences are obviously not reliable statistically between the first and the second generalization stimuli and between the second and the third, even though each was 25 j.n.d.'s removed from the other. Furthermore, and this is much more significant, if we examine the records of individual *S*s, we find: (1) that 19 of the 20 *S*s showed persistent reversals in their CR gradients, that is, they had one or more mean CR's that were greater to more remote than to less remote generalization stimuli; (2) that seven of the 20 *S*s had mean CR's greater to some generalization stimulus than to the conditioned stimulus; (3) that the mean CR's of the one *S* (*S* XI in the table) who had not shown a reversal in his gradient were 23.7, 21.3, 21.1, and 10.2 which, too, manifests little of a gradient between the first and the second generalization stimulus. It is really difficult to see how this experiment can be held up as an exemplar of a lawful CR gradient, and, inasmuch as CR gradients are presumably individual rather than group phenomena, one might even say that Hovland did not find a consistent CR gradient in any of his 20 *S*s, and, at any rate, one might conservatively conclude that wide differences in magnitudes of generalization stimuli produced in Hovland's *S*s only very small and mostly unreliable CR differences. And while there are a few other studies that indicate CR gradients, they, too, have been over-interpreted. Frankly, it might be wise for psychologists engaged in the making of systems to add another postulate, viz., that the numerator of what is known must not be too small a fraction of the denominator of what is unknown.

Hovland's *S*s certainly were conscious of the stimuli for the GSR, and consciousness of stimuli certainly affects the GSR.

Summary

1. Salivary CR's were formed in eight adult human *S*s to 12 different English words.

2. After the CR's were well established, they generalized, in varying amounts, to 32 English words related to the conditioned words semantically, phonetographically, or both—the generalization being greatest from 'take' to 'took,' 72.6 percent, and smallest from 'mock' to 'match,' 8.1 percent.

3. A crude gradient manifested itself in the generalization to the phonetographically related words, the generalization being crudely a function of the number of common phonemes and common letters between the generalization and the conditioned words.

4. In the semantic categories (synonyms were not used in this experiment), the generalization was greatest to contrasts, coordinates, and subordinates of the conditioned words.

5. The generalization to supraordinates of conditioned words was only about one-half of the generalization to the subordinates of these words, and there was surprisingly little generalization from the first to the second word in a compound of two words (*Yankee* Doodle, *mineral* water).

6. The discrepancy between the last finding and what is generally known about habit and association has led to the formulation of a 'subsequent testing situation' hypothesis of CR generalization, according to which, CR generalization develops not during the original training to the conditioned stimuli, but during the subsequent testing for the gen-

eralization stimuli, and the direction of the CR generalization is thus determined not by the relatedness of the generalization stimuli to the conditioned stimuli, but by the relatedness of the conditioned stimuli to the generalization stimuli.

7. The 'subsequent testing situation' hypothesis of CR generalization accounts well for the results of the present study inasmuch as, in accordance with it, the supraordinates of the study were really subordinates and the subordinates were supraordinates, and inasmuch as the presentation of the second word of a compound of two words as a generalization word needed the invoking of backward association to be effective. The hypothesis has also a law of parsimony advantage over the 'original CR training' hypothesis inasmuch as it need not make extra assumptions about the 'spread' or 'irradiation' of CR generalization effects.

8. The 'subsequent testing situation' hypothesis of CR generalization was specifically tested by a supplementary experiment in which two *S*s were conditioned to the second words of compounds of two words (*Doodle*, *water*) and the first words (Yankee, mineral) were presented as generalization words. The results bore out fully the hypothesis, as the generalization CR's were now 57 and 49 percent, respectively.

9. A view is expressed that so-called primary stimulus generalization cannot be considered as something that has been definitely established either in the Russian laboratories or in any laboratory in this country.

(Manuscript received
December 8, 1948)

References

1. Bekhterev, V. M. *Collective reflexology.* Moscow: GIZ, 1923.
2. Beritoff, J. S. *Individually-acquired activity of the central nervous system.* Tbilisi: GIZ, 1934.
3. Cofer, C. N., & Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, **49**, 513–540.
4. Cofer, C. N., et al. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.
5. Doob, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.
6. Foley, J. P., Jr., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
7. Guthrie, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1938, **37**, 412–428.
8. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
9. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
10. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
11. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
12. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
13. Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
14. Hull, C. L., et al. *Mathematico-deductive theory of rote learning.* New Haven: Yale University Press, 1940.
15. Jung, C. G. *Studies in word association.* London: Heinemann, 1918.
16. Keller, M. Mediated generalization: the generalization of a conditioned galvanic skin response established to a pictured object. *Amer. J. Psychol.*, 1943, **56**, 438–448.
17. Kent, G. H., & Rosanoff, A. J. A study of association in insanity. *Amer. J. Insan.*, 1910, **67**, 37–96; 317–390.
18. Konorski, J. *The conditioned reflex and neuron organization.* Cambridge: Cambridge Univ. Press, 1948.
19. Lashley, K. S. An examination of the "continuity theory" as applied to dis-

criminative learning. *J. gen. Psychol.*, 1942, **26**, 241–265.

20. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
21. LOUCKS, R. B. An appraisal of Pavlov's systematization of behavior from the experimental standpoint. *J. comp. Psychol.*, 1933, **15**, 1–45.
22. LOUCKS, R. B. Reflexology and the psychobiological approach. *Psychol. Rev.*, 1937, **44**, 320–328.
23. MAY, M. A. The mechanism of controlled association. *Arch. Psychol.*, 1917, No. 39.
24. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
25. MURPHY, G. An experimental study of the literary versus the scientific types. *Amer. J. Psychol.*, 1917, **28**, 238–263.
26. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford, 1927.
27. PAVLOV, I. P. *Lectures on conditioned reflexes.* New York: International, 1928.
28. PAVLOV, I. P. *Conditioned reflexes and psychiatry.* New York: International, 1941.
29. RAZRAN, G. Theory of conditioning and related phenomena. *Psychol. Rev.*, 1930, **37**, 25–43.
30. RAZRAN, G. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol.*, 1935, No. 191.
31. RAZRAN, G. Salivating and thinking in different languages. *J. Psychol.*, 1935–1936, **1**, 145–151.
32. RAZRAN, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
33. RAZRAN, G. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.
34. RAZRAN, G. Studies in configural conditioning: V. Generalization and transposition. *J. genet. Psychol.*, 1940, **56**, 3–11.
35. RIESS, B. F. Semantic conditioning involving the GSR. *J. exp. Psychol.*, 1940, **26**, 238–240.
36. RIESS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.
37. SHEVALEV, E. *The process of differentiation of the associative motor reflex to tactile stimuli.* Thesis, St. Petersburg.
38. SPENCE, K. W. The nature of discriminative learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
39. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
40. SPENCE, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.
41. WELLS, G. L. A preliminary note on the categories of association reactions. *Psychol. Rev.*, 1911, **18**, 229–233.
42. WOODROW, H., & LOWELL, G. Children's association frequency tables. *Psychol. Mon.*, 1916, **22**, No. 97.
43. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.
44. WOODWORTH, R. S., & WELLS, F. L. The association test. *Psychol. Mon.*, 1911, **13**, No. 57.
45. WYLIE, R. *Generalization of semantic conditioning of the g.s.r.* Unpublished M.S. thesis, Univ. Pittsburgh, 1940.
46. ZALUZHNY, A. S. *Collective behavior.* Kharkov: GIZ, 1929.

STUDIES IN SPATIAL LEARNING: VII. PLACE AND RESPONSE LEARNING UNDER DIFFERENT DEGREES OF MOTIVATION

BY EDWARD C. TOLMAN AND HENRY GLEITMAN

University of California

Introduction

Three previous acticles in this series, two by Tolman, Ritchie, and Kalish (**4, 5**) and one by Ritchie (**2**), seem to have demonstrated that in general it is easier for rats to learn to make two different responses in order to reach the same place than it is for rats to make the same response in order to reach two different places. In the terms of these authors, 'place learning' seems, generally speaking, to be easier than 'response learning.'

The primary purpose of the present experiment was to find out if this initially found difference between place learning and response learning would hold up equally well under both high and low degrees of motivation. A second purpose of the present experiment was to attempt a further analysis of the nature of response learning. It was suggested by Ritchie (**2**) that response learning may really be only a case of double-place learning; the animal learns that, from one starting point in the room, goal place A is the spot to go to; whereas, from another starting point in the room goal place B is the spot to go to. The greater complexity of this double-place learning task would then account for the greater difficulty of what we have called response learning. If this were actually the explanation, then it might be expected that some deprivation of cues or of the sensory apparatus of the animal which permits it to recognize place differences would seriously interfere with response performance. On the other hand, if response learning were 'genuine,' and not just a more complex form of place learning, such sensory impairment would not be expected to disrupt the performances of the response groups radically. Therefore, in the present experiment we tested the relative effects of blinding both the response and the place groups after the original learning had been achieved. This was to see whether the performances of the response groups would or would not be as much disrupted by such blinding as were those of the place groups.

Animals

A total of 32 male M by M pigmented rats from the Tryon 'bright' and 'dull' stocks were used for this experiment. All were approximately 100 days old at the beginning of the experiment, were experimentally naive, and had been on a 24-hour wet food schedule for one week prior to the experiment. During the experiment they were divided into four groups of eight each as follows: IA and IB—place learners under 12 hours and 46 hours of food deprivation, respectively, and IIA and IIB—response learners also under these same two respective degrees of food deprivation. All groups were run every other day from about 12 midnight to 4 A.M.

Apparatus

The apparatus was an elevated T-maze, built from 2-in. wide, black wooden strips. The center path connecting the two goal boxes G-1 and G-2 (see Fig. 1) was 12'6" long, path AB was 3'7". The only illumination in the room came from a standing lamp with a one-foot diameter reflector and a 150-watt bulb, standing about two feet above the level of the maze runways at the position indicated in the figure. The lamp faced the left wall, the approximate angle to incidence with the wall being about 30 degrees.

At the two ends of the center path were two goal boxes, 8" long, 4" high, and 4½" wide, each

containing a metal cup which could be filled with wet food.

The maze was set in a room 18 × 15 feet. The home cages were kept in a rack about five feet behind point A.

Procedure

Preliminary training.—All groups received the same preliminary training, under 24 hours of food deprivation. During this training the animals were run in a different room from that used in the experiment proper. This preliminary training consisted in running the animals on a six-foot elevated straight-away with goal-boxes containing wet food at one end. The animals were run for five days on this set-up, each rat being given four trials per night.

Experimental procedure—Pre-operative.—All rats were run on the maze illustrated in Fig. 1,

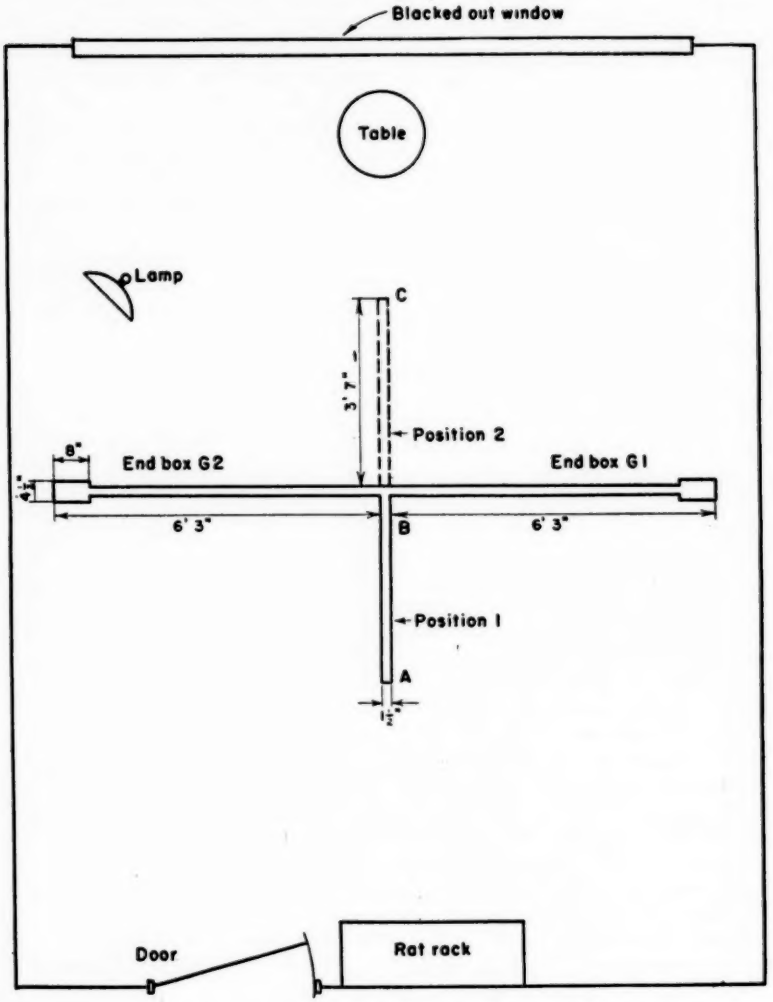


Fig. 1. Floor plan of maze in its environment

the 12-hour deprived groups being run on alternate nights from the 46-hour deprived groups. Each animal received six trials per night. On three of the six trials the maze was in the position shown in Fig. 1 (Position 1) and on the other half the entire maze was rotated 180 degrees, so that the starting point would be at point C (Position 2). The order of these positions was 112212 and 221121 on alternate experimental nights.

Animals in the two place learning groups, IA and IB, always found food in the same 'place,' regardless of where they started from, i.e., they either found food always in the goal-box on the illuminated side of the room, or in the goal-box in the shadow. For half of the animals in each group, the illuminated side was rewarded, and for the other half of the animals the side in shadow was rewarded. The animals were thus required to learn that food was at a particular location in the room. The cues upon which they depended were presumably visual and derived from the position of the light.

The animals in the two response learning groups, IIA and IIB, always found food in the same direction relative to their starting point, so that sometimes it would be in the shadow and sometimes in the light, but it would always require the same turn (right or left) at the choice point. For half of these animals a left turn was the correct response, and for half a right turn. These animals were thus required to learn to make a particular 'response' (e.g., a right or left turn) to get to the food, regardless of illumination.

All groups were given six trials a night by the 'correction method' for seven nights. (A rest period for all groups of one extra 24-hour period was inserted between the third and fourth experimental nights, owing to a too severe weight drop on the part of groups IB and IIB.) Errors, times, and number of VTE's were recorded.

Experimental procedure—Operative and post-operative.—In order to test the hypothesis that response learning may be but a complex form of place learning, all animals were blinded on the night following their last performance on the maze, on the assumption that, under the conditions of this experiment, the main cues for place learning were visual and that therefore loss of vision should interfere with any performance based primarily on place learning. Blinding was effected by removing the eye-balls under ether. One animal in Group IIB died as a result of the operation.

The animals were given one 24-hour period with full food and water supply as a recovery period, before continuing with the experiment. On the night following, they were run again on the same set-up with the same degree of motivation and in the same manner as before the operation. This was continued for four days.

Results

Pre-operative.—The mean number of errors per rat, and mean VTE's per rat for the seven nights prior to the operation for the four groups are shown in Table I. These figures indicate that the place groups learned more readily than the response groups and that increased motivation (at least within the limits employed here) facilitated better performance for both types of set-up; and a similar pattern held for the VTE's. (The time data were not worked up but also presented a similar picture.)

TABLE I

Mean Numbers of Errors and VTE's per Rat During Pre-Operative Training

	Mean Errors (N =8)	Mean VTE's (N =8)
IA 'Place,' 12 hr. deprivation	12.5	55.6
IB 'Place,' 46 hr. deprivation	3.9	24.5
IIA 'Response,' 12 hr. deprivation	19.7	86.7
IIB 'Response,' 46 hr. deprivation	11.3	41.1

The same relationships are presented as a function of trials in Fig. 2 and Fig. 3. The curves also indicate that, generally speaking, the mean VTE's and the mean errors per day go up and down together.

Post-operative.—Fig. 2 shows also that, after the operation, the mean number of errors on the first day of testing went up for all groups except IIA (Response Group, 12 hr.). On the fourth day after the operation both response groups had reached a level of performance better than the one attained on the last night before the operation. The two place groups, while showing some improvement during this period, were still consider-

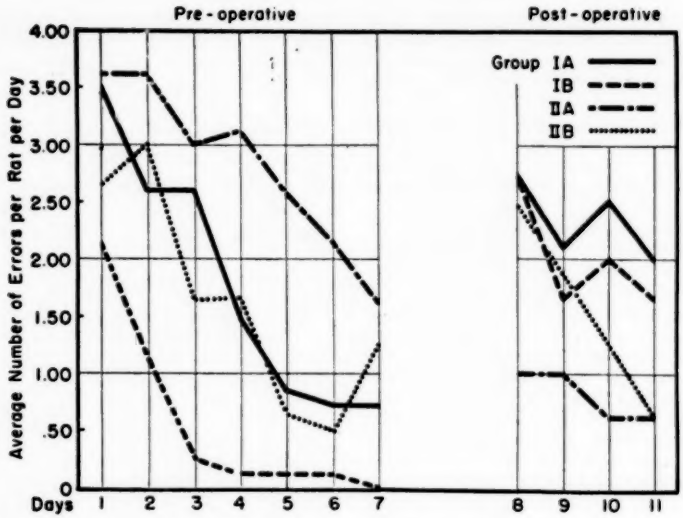


FIG. 2. Average number of errors per day per rat for all groups
(Group IA —'Place' —12 hours deprivation
Group IB —'Place' —46 hours deprivation
Group IIA—'Response'—12 hours deprivation
Group IIB—'Response'—46 hours deprivation)

ably poorer than they had been on the last night before the operation.

It is to be noted also that there were considerable individual differences among the operated place learners. While the majority of these animals made many errors and seemed very hesitant and confused, about one-fourth of them (one in Group IA and 3 in IB) learned very readily, making two or less errors on the last three nights of running out of a total possible 18. Presumably these non-upset animals utilized some non-visual extra-maze cues, very probably auditory ones.

As shown in Fig. 3, the number of VTE's after operation increased for all groups, and decreased somewhat after four days of running. The VTE's for Group IIA went up more, however, than those for IIB.

DISCUSSION

According to sign-Gestalt theory, learning in a maze is conceived as consisting in the building up of field expectancies. These field expectancies may be formed primarily in terms of particular places or directions (as defined in terms of lights, noises, etc.) or in terms of particular responses (as a left body turn, a right body turn, etc.). In the two types of maze set-up used in the present experiment, the so-called 'place groups' presumably tended to build up place or direction expectancies whereas the so-called 'response' groups presumably tended to build up 'response' expectancies.

In any event, the earlier findings of Tolman, Ritchie and Kalish (4, 5) were confirmed, to wit: that the acquiring of place performance is easier

on such a type of set-up with strong directional cues than is the acquiring of response performance. This was also found to hold for each of the two degrees of motivation which we used.

It was earlier suggested by Ritchie (2) that the difference in relative difficulty in acquiring the two kinds of performance may be due to the fact that the response performance may really depend not on response expectancies but on a more complicated double place expectancy—namely, that of learning to go to one place from one starting point, and to another place from another starting point. If this latter interpretation were correct, and if the animals use primarily visual cues for their place expectancies, we should expect that response learners (and naturally place learners as well) would become severely handicapped by a loss of vision. Our results indicated that the place performers were severely handicapped except for a few individuals (who presumably used auditory cues). For the response performers the results were not altogether clear-cut. One subgroup (IIA) showed no sign of disturbance, whereas the other subgroup (IIB) was somewhat affected initially. However, the outcomes for these response groups were nevertheless quite different from what would have been expected if response learning were a complicated form of place learning. For both response groups relearned quickly and far excelled the place groups after the four days of such relearning. This would indicate that

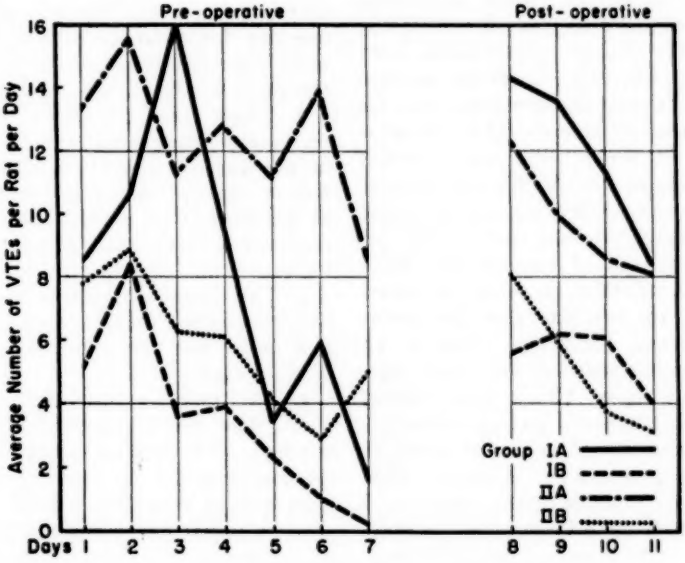


FIG. 3. Average number of VTE's per rat per day for all groups
Group IA —'Place' —12 hours deprivation
Group IB —'Place' —46 hours deprivation
Group IIA—'Response'—12 hours deprivation
Group IIB—'Response'—46 hours deprivation

practically no place expectancies (involving vision) were involved in the performance of these response groups.

Consider now the VTE's. As already remarked, the VTE curves followed, on the whole, the error curves. This fits in with previous findings concerning VTE and errors on a maze (Jackson, 1) and the general conclusions of Jackson (3) and of Tolman and Ritchie (3) that, in a maze, VTE's are analogous to partial errors and hence are positively related to errors (rather than negatively as in the usual discrimination set-up). This fact of the positive correlation between VTE's and errors is to be seen (a) in the fact that the place groups (which made fewer errors) showed fewer VTE's than the response groups, and (b) in the fact that in general the more highly motivated groups (which also made fewer errors) also made fewer VTE's. It is to be observed furthermore, however, that there was a general increase in VTE's after the operation even for that group of animals (IIA) the error scores of which were not affected. This suggests that VTE's also increase merely as a result of an unaccustomed bewildering situation such as arose in the new task of running the maze blinded. Finally, it is to be noted that on the first day after the operation Group IA (Place learners, 12 hours food deprivation) made considerably more VTE's than Group IB (Place learners, 46 hours deprivation) even though the error scores of the two groups were identical. This suggests a possible direct influence of motivation on VTE's: namely, that, for one and the same degree of knowledge (errors) higher motivation, because of the increase in urgency involved in it, may, as such, reduce the number of VTE's.

Summary

1. Four groups of rats, eight in each group, were run on a Tolman, Ritchie and Kalish (4) T-maze with two opposite and alternative starting paths. One main group was run to the same food place (*place group*) from the two starting positions. The other main group (*response group*) ran to two different food places from the two different starting positions in such a way that they always had to make the same body turn, right or left, in order to reach the food. Each of these two groups was further subdivided into a highly motivated (46 hour deprivation) group and a low motivated (12 hour deprivation) group. All four groups were run by the 'correction' method.

2. After seven days (six trials per day) of initial learning the animals in all four groups were blinded and after one day for recovery were run four more days (six trials per day) under the same conditions as previously.

3. Under the conditions of directional lighting used both here and in the earlier Tolman, Ritchie and Kalish experiments (4, 5) we found, as did they, that place performance is acquired faster than response performance.

4. This faster acquisition of place performance was found to hold equally well under our two different degrees of motivation.

5. The place groups were more upset than the response groups by blinding. This indicates that under the conditions of our experiment the place groups depended for the most part on visual cues, whereas the response groups for the most part did not. It appears therefore improbable that the response groups had come to perform upon the basis of double place expectancies.

6. Our findings as to the relations between VTE's and errors confirmed the earlier findings of Jackson (1) and the explanations of Jackson (1) and of Tolman and Ritchie (3) that on a maze VTE's are, by and large, equivalent to partial errors.

7. However, the fact that although on the first day after the operation the number of VTE's of the 12-hour response group was greater than that of the 46-hour response group (although the mean errors for these two groups were the same) suggests that, in addition, high motivation, as such, may decrease VTE's, irrespective of knowledge (errors).

(Manuscript received
December 14, 1948)

References

1. Jackson, L. L. VTE on an elevated maze. *J. comp. Psychol.*, 1943, 36, 99–107.
2. Ritchie, B. F. Studies in spatial learning: III. Two paths to the same location and two paths to two different locations. *J. exp. Psychol.*, 1947, 37, 25–38.
3. Tolman, E. C., & Ritchie, B. F. Correlation between VTE's on a maze and a visual discrimination apparatus. *J. comp. Psychol.*, 1943, 36, 91–98.
4. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: II. Place learning vs. response learning. *J. exp. Psychol.*, 1946, 36, 221–229.
5. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning: V. Response learning vs. place learning by the non-correction method. *J. exp. Psychol.*, 1947, 37, 285–292.

SIGN LEARNING AS A FACTOR IN EXTINCTION [1]

BY JOHN P. SEWARD AND NISSIM LEVY

University of California at Los Angeles

Problem

Razran's (21) thorough analysis makes any introductory survey of theories of extinction unnecessary. For the purpose of classification Hilgard and Marquis's (8) twofold division of inhibitory processes remains as convenient as any. They distinguish between *adaptation*, "the reduction in response resulting from continued elicitation of the response itself," and *interference*, "the reduction resulting from the simultaneous elicitation of another incompatible response" (p. 105). The various theories cited by Razran may be grouped roughly in these two categories. Adaptation includes, for example, Pavlov's concept of internal inhibition and Hull's (10) 'reactive inhibition.' As defined above it also includes Skinner's (22) treatment of extinction as the exhaustion of a 'reflex reserve,' although Skinner scrupulously avoids the term inhibition. Interference includes Guthrie's (7) 'inhibitory conditioning,' Wendt's (26) 'competition between behavior systems,' and Hull's (10) 'conditioned inhibition,' among others.

It is quite likely that neither process by itself is the sole basis of extinction. But there is a danger that we may underestimate the importance of interference because of a weakness of most theories of this class; namely, they provide no method of eliciting the incompatible response in the first place. Before such a response can interfere it must be made, and before it can be made the original response must be prevented. True, if the CS is combined with a stronger stimulus to an opposing response, as when Jones (15) combined the feared rabbit with the appearance of appetizing food, an interference theory can be directly applied. But most laboratory CRs are extinguished by merely withholding food or shock. For this reason Guthrie (7) had to supplement his theory by other factors making for response decrement; e.g., sensory adaptation, fatigue, and refractory period. For the same reason Mowrer and Jones (19) and Miller and Dollard (17) presupposed a 'fatigue drive' before the resulting cessation of response could be conditioned. Hull's concept of reactive inhibition serves a similar purpose.

Another alternative lies in the type of theory sponsored by Tolman (24) and Culler (2). The former, following Williams (27), incorporates extinction in his sign-learning theory of conditioning. "The CR is a response to the sign-relationship $S_2 \rightarrow S_1$. And this means that it goes off only so long as that sign-relationship is 'believed' in by the O" (24, p. 330). In fact, Tolman uses extinction as evidence of this contingency of CR on the sign character of CS. Culler goes one step further in identifying extinction with 'inverse conditioning.' "Without shock the bell becomes a *safety*-signal even as with shock it was a *danger*-signal" (2, p. 150). According to these formulations conditioning depends on two things: the sign function of CS and the incentive function of US. Extinction depends on the same functions operating under altered conditions. As ordinarily brought about by simply removing

[1] Aided by a grant from the Committee on Research, University of California.

US, it follows because the second function is lost. Even the earliest stage of response decrement can thus be accounted for.

A suitable name for the Tolman-Culler hypothesis presents some difficulties. It explains extinction as due to a loss of motivation from an external source. The term *demotivation*, however, would imply a loss of drive, or satiation, which in the case of a hungry animal seeking food would be obviously untrue. On the other hand the term *decathexis*, suggested to the senior author by Dr. David Krech, would not apply to Culler's example of shock avoidance, since here it is the secondary drive of fear that is removed rather than the incentive of escape. Since in either case the argument goes back to the signalizing property of CS, it is appropriate to call the Tolman-Culler hypothesis the *sign-learning theory of extinction.*

How can this theory be tested? Fortunately, it leads to one deduction that adaptation theories do not: namely, that a response can be extinguished without being made. As Spence (23) has pointed out, sign learning is essentially an S—S rather than an S—R process. If extinction is such a process, then, contrary to the definition of adaptation quoted above, it should be feasible without evoking R. Suppose an organism has learned a response preparatory to entering a goal place and finding food. To extinguish such a response the sign-learning theory demands merely that he discover and recall that the goal place is now empty. Its requirements are largely fulfilled by giving him ample opportunity to explore the goal situation; nothing is said about making the critical response before getting there. The present experiment is therefore designed to provide such a sequence of experience and to measure its effect on the rate of extinction.

Method

The response selected for conditioning and extinction was the operant introduced by Graham and Gagne (6); namely, that of traveling an elevated pathway from a starting platform to a goal platform. The pathway consisted of a strip of redwood 36 in. long by 1¾ in. wide. At each end was a plywood platform; the starting platform (A) was 10 in. by 8 in., the goal platform (B) was 10 in. by 4 in. Both platforms were separated from the pathway by transparent sliding doors. At the far end of platform B a small foodcup was mounted in a block so that before reaching it a rat could not see if it contained food or not. The apparatus was unpainted and stood three feet from the floor in one corner of a long, narrow room. No attempt was made to provide a uniform surrounding. A third platform (X), 10 in. by 6 in., of bare plywood, stood by itself about six feet away.

The population consisted of 33 albino rats, 13 females and 20 males, from 4½ to 5½ months old at the time of training. They were maintained during the experiment on a Steenbock diet in the form of a wet mash fed right after the daily run for 30 min., except for the three days before extinction, when the time was cut to 20 min.

The procedure falls naturally into four parts as follows:

1. *Adaptation.*—This period lasted from eight to 13 days for different animals, during which they were given about 15 minutes a day of habituation to elevated pathways. On the last three days they were placed on the apparatus itself.[2]

2. *Training.*—The day after the last adaptation each rat was given 10 training trials with food at B. The rats were run four at a time (except for one group of five) and took their trials in rotation. A trial started with the door at A closed, the one at B open. *E* placed the rat on A facing to the rear, counted to three, and raised the door. When the rat had entered the farther door *E* lowered it behind him, allowed

[2] For 26 rats the kind and amount of exploration permitted during these three days were systematically varied for another purpose. Since the variations were equally apportioned between the experimental and control groups of this experiment and since they produced no significant differences in the results of training, they will not be described here.

TABLE I

TIME SPENT ON PLATFORM B OR X DURING PRE-EXTINCTION AND EXTINCTION

Group	Pre-extinction	Extinction	
		Before Trials	Between Trials
E	5(120 sec. on B)	120 sec. on B	120 sec. on B
C	5(120 sec. on X)	120 sec. on X	20 sec. on B 100 sec. on X

him to eat for 20 sec., then returned him to his cage to await the next trial. By using two stopwatches *E* obtained two time scores: (1) *latent time*, measured from the raising of the door till all four feet were on the pathway; and (2) *running time*, measured from the end of latent time until all four feet were on B.

3. *Pre-extinction.*—At this point the rats of each batch were divided between two groups, an experimental group (E) of 17 (six females and 11 males) and a control group (C) of 16 (seven females and nine males). The day after training, each E rat was placed directly on B with the door down and the foodcup empty and left there for two min. Each C rat spent two min. on platform X. Five such treatments were given in rotation.

4. *Extinction.*—On the following day all rats were given extinction trials to a criterion of two successive refusals to leave platform A within three min. Individual extinction trials were identical with training trials save for the absence of food on B. Rotation of rats, however, was now abandoned. Each E rat was first placed on B without food and confined there for two min., then removed to A for trial 1. The two-min. interval between trials was also spent on B. Each C rat, on the other hand, spent the first two min. on X; between trials, after 20 sec. on B he was placed on X for 100 sec. Table I summarizes the pre-extinction and extinction treatments of the two groups.

It will be seen that the sole difference between the experimental and control conditions was that group E spent more time in the goal situation without food. If sign learning is a factor in extinction we should therefore expect group E to extinguish more rapidly than group C.

RESULTS

Training scores.—Before comparing the performance of the two groups in extinction we must be sure that any difference found cannot be attributed to a difference in strength of conditioning. Two measures of response strength were obtained in the course

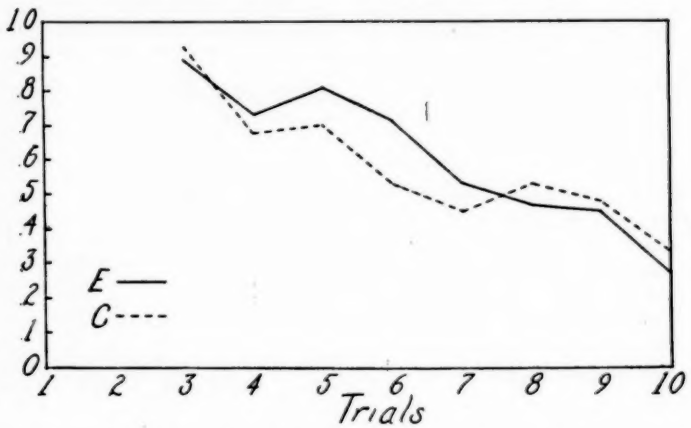


FIG. 1. Mean $\log_{10}$ latent times on training trials of experimental and control groups

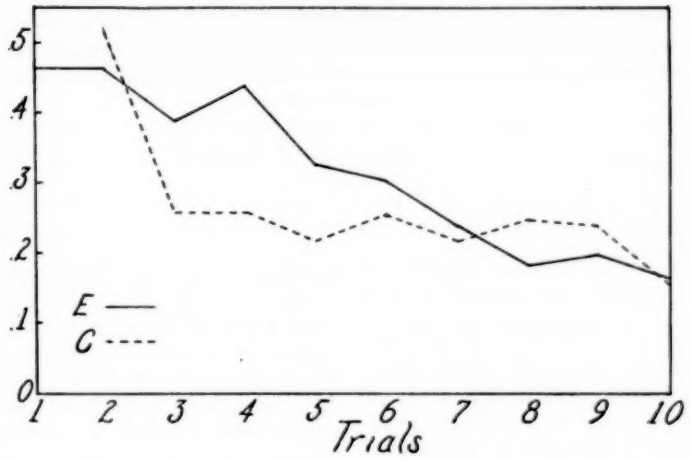


FIG. 2. Mean $\log_{10}$ running times on training trials of experimental and control groups

of training, as described in detail above: (1) latent time, or the time taken to leave platform A, and (2) running time, or the remaining interval before arriving on B. Because these measures were badly skewed toward the longer times, we followed Graham and Gagne in converting them into logarithms. Fig. 1 shows the mean log latent times of the experimental and control groups in the last eight training trials. (The first two trials were omitted because several rats failed to start in three min. and had to be urged.) Fig. 2 presents mean log running times.

Two tests of the null hypothesis were applied to each measure:

1. Subtracting the mean log time of group C from that of group E on each trial, we applied the *t*-test to the mean of the differences. The results were as follows: for latent times on trials 3 to 10, $t = 0.9$, $P = .4$; for running times on trials 2 to 10, $t = 1.31$, $P = .3$.

2. Individual mean log times on the last four training trials were obtained and averaged for each group; *t* was then found for the difference between the group means. The results appear in Table II. Although *t* was calculated from the logarithms, the group mean logs are here converted back into mean times for greater ease of interpretation; since these are derived from the distribution of logs they are the geometric means of the original time scores. None of these statistics makes it unreasonable to assume that the learning scores of groups E and C were samples drawn from a single population.

Extinction scores.—a. Number of trials to criterion: Table III gives frequency distributions of the number of trials to extinction, not including the criterion trials, for the two groups.

TABLE II

GEOMETRIC MEANS OF LATENT TIMES AND RUNNING TIMES (IN SEC.) ON TRIALS 7–10 OF TRAINING

	Group E	Group C	C-E	*t*	*P*
Latent time	2.78	2.80	.02	.02	>.9
Running time	1.63	1.70	.07	.49	.7

TABLE III

FREQUENCY DISTRIBUTION OF TRIALS TO REACH CRITERION OF EXTINCTION

Group	Number of Trials															
	0	1	2	3	4	5	6	7	8	9	10	11	12	23	24	25
E	1	4	4	3	1	2		1				1				
C		2	1	1		1	4		2	1	1		1	1		1

It will be noted that one member of group E reached the criterion in zero trials; i.e., he refused to cross at all. One other rat in the same group failed to leave the platform within three min. on trial 1, although he did cross once on trial 2. Fig. 3 presents the same data expressed as cumulative percentages. The means appear in Table IV, where it will be seen that the difference of 5.13 trials is significant at the .01 confidence level.

b. Time scores on trial 1: The fact remains that most of the experimental group, even after their pre-extinction experience on B, crossed one or more times to the empty platform. The question arose: did these rats take longer to cross than the controls? Trial 1 was the most interesting, since it was uncontaminated by any effects of previous unrewarded responses. Moreover, the number of unextinguished rats available for comparison

TABLE IV

MEAN NUMBER OF TRIALS TO REACH CRITERION OF EXTINCTION

Group	*N*	*M*	C-E	*t*	*P*
E	17	3.12	5.13	2.9	.01
C	16	8.25			

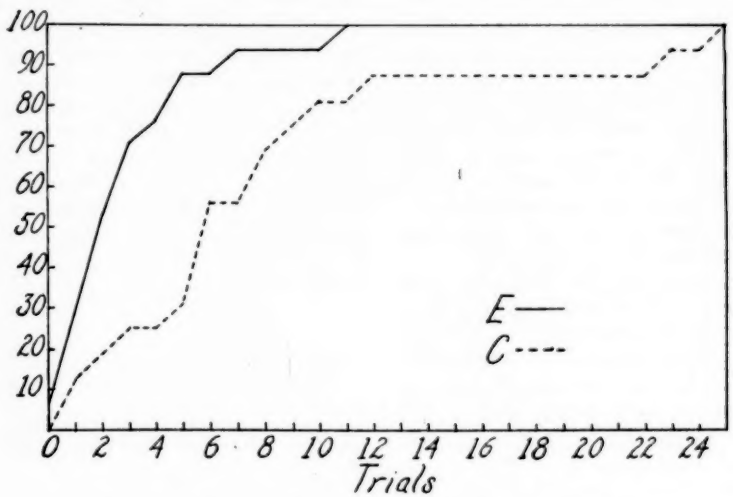


FIG. 3. Percentages of experimental and control groups reaching criterion in a given number of trials or less

dwindled rapidly thereafter. The data for trial 1 appear in Table V. These data are comparable with those of Table II, covering the last four training trials, except for the omission of the two E rats that failed to cross. It will be seen that in group C both latent time and running time showed an increase over their previous values. In group E latent time increased by a similar amount, but running time was prolonged to a value three times as great as that of C, producing a difference significant at the .02 level. That is, the E rats that left platform A did so as readily, on the average, as the C rats, but it took them considerably longer to reach B. These figures bear out the strong impression of the *E* who observed their behavior in the early extinction trials. It was characteristic of them to start off with little hesitation but to stop abruptly at the end of the path and enter the door to B only with great reluctance. Evidently their pre-extinction treatment exerted a stronger effect on those parts of the habit sequence close to the goal than on earlier components.

It is of interest to know whether the effect in question was simply to undo the results of training or to cause some other change of behavior. Qualitatively the behavior of the *E* rats was quite different from what it had been before training. To get quantitative confirmation we selected the 15 E rats represented in Table V and computed their geometric mean running time on the first training trial. On that trial, before ever finding food on platform B, their time was 2.62 sec., while on their first extinction trial, after finding *no* food on B, they crossed in 7.70 sec. The difference is significant at close to the .02 level. It strongly suggests that the result of pre-extinction treatment cannot be described as a mere loss of the effects of training.[3]

TABLE V

GEOMETRIC MEANS OF LATENT TIMES AND RUNNING TIMES (IN SEC.) ON TRIAL 1 OF EXTINCTION

	Group E $N=15$	Group C $N=16$	E-C	t	P
Latent time	7.30	6.22	1.08	0.45	.7
Running time	7.70	2.55	5.15	2.65	.02

DISCUSSION

According to Skinner's "law of extinction of chained reflexes" (22, p. 105), only members actually elicited undergo extinction. He was talking about a situation in which the chain was broken somewhere in the middle and only the earlier reflexes were elicited. The present experiment shows that if the last member is elicited without the earlier ones the entire chain is weakened. This finding demands not only a revision of the law but a search for other principles of explanation.

Adaptation theories are inadequate to account for the more rapid extinction of the experimental group, since by definition they presuppose the elicitation of the response. In our experiment the critical response was not elicited in the pre-extinction period; therefore it could not have been directly weakened. The alternative is some form of interference theory.

But what interfered with what? On the level of overt behavior the response of stopping, or not starting,

[3] It is fortunate that this is the case, since otherwise it would have been impossible to demonstrate any quantitative effect by means of this measure. The reason is that the 13 control rats that crossed within three min. on their first training trial gave a geometric mean running time of 2.40 sec. as compared with 2.56 sec. on their first extinction trial, an insignificant difference.

interfered with that of going ahead. Such a description is hardly explanatory. A more satisfactory answer is available in sign-learning, or expectancy, theory. A free rendering of Tolman might run something like this: During training the rats started sooner and ran faster largely because they developed a hypothesis that food was to be found on platform B. Before and between extinction trials the E rats had more time than the C rats to develop the contrary hypothesis. Essentially, then, it was the interference of one expectation with another that led to quicker extinction.

This is not to imply, of course, that sign-learning theory explains all the facts of extinction. In our own experiment the procedures used accounted for only a part of the total process, leaving plenty of room for other factors to operate. The relative importance of these factors is a question raised but not answered by our data. All the present experiment does is to point to expectancy as *one* factor that must be reckoned with.

A second question raised by our results is, what does this extinctive process do? Specifically, does nonreward expectancy operate simply by subtraction; i.e., does it merely remove one vital factor, the incentive, from the situation motivating the animal? Or, by blocking the incentive factor, does it produce an 'emotion' which actively opposes the animal's forward progress? That something of the latter sort happens is suggested by the fact noted above; namely, that the E rats were slowed up not only in comparison with their final training trial but with their first one. Evidence in the same direction has been reported by Miller and Stevenson (18), Skinner (22), and Crespi (1).

Through their agreement with a prediction from sign-learning theory our findings lend some measure of support to that theory. One way of looking at them is in relation to latent learning. In the latent learning experiment a response suddenly appears that could not be predicted on the basis of previous 'reinforcement.' The present experiment may be said to demonstrate 'latent extinction,' in the sense that a response decrement appeared that could not be predicted from prior 'non-reinforcement.'

Again, our data have some bearing on the problem of partial reinforcement. Humphreys (11, 12, 13) found that human *Ss* reinforced on only half their training trials extinguished more slowly than after reinforcement on all. He fitted his discovery to an expectancy hypothesis. Skinner (22) demonstrated a similar phenomenon in rats by means of 'periodic reconditioning.' But further experiments (3, 4, 5, 14, 20) have thrown considerable doubt on Humphreys' interpretation as applied to rat behavior. If expectancy has not been shown to *retard* extinction at this level, the present findings indicate that it may *accelerate*—and thus play some part in—the process.

It should be pointed out, however, that a sign-learning interpretation does not rule out other theoretical approaches. Hull (9), for example, in 1931 discussed the disruption of a habit on removal of reward in terms of the loss of r_0, the fractional antedating goal reaction, from the stimulus complex. In *Principles of behavior* (10, pp. 100f., 131) he presented a similar argument, and predicted that removal or decrease of primary reinforcement would lead to "a progressive weakening of the habit," without, however, making explicit the grounds for that prediction. Again, in 1933

Tolman (25) set up a test of sign-learning theory very similar to our own, but involving punishment instead of extinction. Two years later Miller (16) derived *and verified* the same prediction from S—R reinforcement principles. It is altogether conceivable that the present experiment would yield to a similar treatment. One of our findings in particular—that running time was more affected than latent time—should lend itself readily to an interpretation in terms of stimulus generalization.

But in their present stage of development distinct theories serve a useful purpose by setting up divergent predictions for experimental test. What would happen, for example, if we greatly increased the number of training trials given our rats before pre-extinguishing them on the goal platform? A reinforcement theorist would hold that the added strength of the original habit would make it more resistant to the effects of removing reward. The experimental rats should therefore approach the performance of the controls. A field theorist like Lewin, on the other hand, would point out that overtraining should increase the unity of the path from platform A to platform B. As a result, on the first extinction trial the experimental animals might be expected to delay not only their approach to B but their departure from A.

Eventually the exciting contours of today's theoretical peaks will yield to the erosive influence of fact-finding. When that time comes, 'expectancy' itself may be found to emerge as a higher-order construct from a more primitive set of assumptions. To explore the possibilities in this direction would take us beyond the limits of the present report. It is clear, however, that any such attempt must find an unequivocal place for goal anticipation as a vital factor in learning and for 'non-goal' anticipation as a vital factor in extinction.

Summary

An experiment was designed to test the sign-learning theory of extinction; the theory, namely, that when reward is removed CS becomes a sign of non-reward and CR drops out through lack of incentive. The deduction tested was that CR could be weakened without itself being elicited, provided that the situation formerly associated with reward was now experienced without it.

The response selected was to traverse a narrow elevated path from one platform to another; the reward was found on the second platform. Thirty-three rats were given 10 training trials in this habit. They were then divided into two groups, experimental and control. Group E spent time on the goal platform without food both before and between extinction trials; group C spent equivalent periods on a neutral platform. Both groups were extinguished to a criterion. The results were as follows:

1. Group E reached the criterion in a mean of 3.12 trials, group C in 8.25; the difference was significant at the .01 level.

2. Fifteen of the 17 rats of group E crossed on the first extinction trial. Their starting (latent) time was comparable with that of group C, but their crossing (running) time was prolonged by an amount significant at the .02 level.

3. Group E's running time on the first extinction trial was significantly slower than on the first training trial; the corresponding difference in group C was insignificant.

These results are taken to verify the hypothesis that led to their prediction, with one qualification: the learning that underlies extinction may serve not merely to subtract the incentive but to produce an 'emotional' interference with the response.

(Manuscript received
November 16, 1948)

References

1. Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.
2. Culler, E. A. Recent advances in some concepts of conditioning. *Psychol. Rev.*, 1938, **45**, 134–153.
3. Denny, M. R. The role of secondary reinforcement in a partial reinforcement situation. *J. exp. Psychol.*, 1946, **36**, 373–389.
4. Finger, F. W. The effect of varying conditions of reinforcement upon a simple running response. *J. exp. Psychol.*, 1942, **30**, 53–68.
5. Finger, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **31**, 120–134.
6. Graham, C. H., & Gagne, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.
7. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
8. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
9. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.
10. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
11. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
12. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
13. Humphreys, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–76.
14. Humphreys, L. G. The strength of a Thorndikian response as a function of the number of practice trials. *J. comp. Psychol.*, 1943, **35**, 101–110.
15. Jones, M. C. A laboratory study of fear: the case of Peter. *Ped. Sem.*, 1924, **31**, 308–315.
16. Miller, N. E. A reply to 'Sign-Gestalt or conditioned reflex?'. *Psychol. Rev.*, 1935, **42**, 280–292.
17. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
18. Miller, N. E., & Stevenson, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.
19. Mowrer, O. H., & Jones, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369–386.
20. Mowrer, O. H., & Jones, H. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.
21. Razran, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.
22. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
23. Spence, K. W. Theoretical interpretations of learning. In: Moss, F. A. (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1942, 280–329.
24. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
25. Tolman, E. C. Sign-Gestalt or conditioned reflex? *Psychol. Rev.*, 1933, **40**, 246–255.
26. Wendt, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–281.
27. Williams, K. A. The conditioned reflex and the sign function in learning. *Psychol. Rev.*, 1929, **36**, 481–497.

RETENTION AND WARMING-UP EFFECTS IN PAIRED-ASSOCIATE LEARNING

BY ARTHUR L. IRION

University of Illinois[1]

The warming-up effect in retention has been defined (3) in terms of the greater slope of the relearning curve as compared with the slope of the original learning curve at a corresponding level of proficiency. This effect was first noted in relation to the problems of learning and retention by Thorndike (7) in 1914 and has, since that time, received relatively scant attention. Bell (2) has noted, in the Koerth pursuit-rotor situation, that amount of warming-up effect (as measured by the difference in score between the first and second relearning trials) is a function of the length of the retention interval. His results suggest that, as length of the retention interval increases, the amount of warming-up effect tends at first to increase and then to decrease. Ammons (1), in his theoretical analysis of the motor learning situation, speaks of amount of warm-up decrement (that is, amount of loss of proficiency attributable to a need for warming up) rather than to amount of warming-up effect. In this analysis, Ammons assumes that amount of warm-up decrement present in relearning is influenced by four experimental variables; total duration of previous practice, duration of practice since the beginning of relearning, the number of previous practice sessions, and the length of the intersession rests. Five assumptions are made concerning warm-up decrement (which Ammons identifies by the symbol, Dwu). These are:

1. Initial post-rest Dwu increases as a negatively accelerated increasing function of total duration of previous practice.

2. Dwu will be a positively accelerated decreasing function of the duration of practice since intersession rest.

3. The 'shape' of the curve of Dwu within a practice session will always be the 'same,' whatever the initial amount of Dwu.

4. Initial post-rest Dwu decreases as a positively accelerated function of the number of previous practice sessions.

5. Initial post-rest Dwu is a negatively accelerated increasing function of the duration of intersession rests.

Some of these empirical relations have received prior attention. Jost's (4) well-known principle ("If two assocations are of equal strength but of different age, a new repetition has a greater value for the older one") may be interpreted either as a statement of the relationship between amount of warming-up effect and initial associational strength (amount of prior practice), or as a statement of the relationship between amount of warming-up effect and the length of the retention interval. In still another form, Luh (6) described the latter relationship when he noted that the retention curve for relearning does not fall as rapidly as does the retention curve for recall. This, of course, implies that the slope of the relearn-

[1] From the laboratories of psychology at the State University of Iowa. The writer wishes to express his indebtedness to Prof. Kenneth W. Spence who directed the investigation. The data reported here were contained in a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree, State University of Iowa, 1947.

ing curve increases as the length of the retention interval increases.

In a recent article (3) the writer has described some possible interpretations of this warming-up effect in terms of variations in the performance set of the subject between the time of original learning and the time of recall. In that article was discussed the possibility of removing a portion of the retention loss which normally occurs over a rest period by allowing the *S* to engage in a warming-up activity before recall. Such a warming-up activity should be similar to the task to be recalled, but, of course, should not provide specific practice in connection with the previously learned task. Some evidence that pre-recall warming-up might be efficacious in producing higher recall performance may be gleaned from the studies on reminiscence, in which it has been shown that differences in the activity which fills the retention interval have an influence on the amount of reminiscence obtained. For example, Ward's (8) study may be cited. Having completed an investigation of reminiscence in the rote-serial learning situation, Ward was concerned that the rest interval activity permitted the *Ss* (light reading) might have offered an opportunity for rehearsal. In order to control this possibility, he ran a supplementary experiment in which the rest interval was filled with a color naming task. This task was one which required a fairly rhythmical response of the *S* and thus, it may be presumed, constituted a task more similar to the rote learning task than was the light reading. Ward's results showed that the *Ss* who engaged in color naming were superior in recall to the *Ss* who engaged in reading. A reminiscence score of 120 percent was obtained with the former group, whereas a score of 104 percent was obtained with the latter, the difference being 4.6 times its probable error. These results suggest that the superiority of the color naming group was due to the fact that the color naming served better to maintain his *Ss* in a warmed-up condition than did the reading.

The present paper is a report of two related experiments. The first is concerned merely with a determination of a retention function for the materials used in these studies and also a determination of whether or not the warming up effect in the verbal learning situation varies as a function of the length of the retention interval. The second experiment is concerned with an attempt to decrease the loss of retention which follows the introduction of a rest interval by causing the *Ss* to engage in warming up activities immediately before the start of relearning.

Procedure

The learning task.—*Ss* were required to learn a 15-unit list of paired, two-syllable adjectives by the method of paired-associates. The pairs of words used in the experiments were as follows: Careful-Upper, Daring-Profuse, Brazen-Musty, Toxic-Noonday, Eager-Crescent, Wayward-Garish, Haughty-Broken, Narrow-Zealous, Graphic-Quiet, Yielding-Festive, Limpid-Innate, Stalwart-Obscure, Unjust-Rounded, Kindred-Artless, Juicy-Verbal. This list was presented in three different orders upon a Hull-type memory drum at a rate of approximately four sec. per pair of words. Calibration of the apparatus revealed that, in presenting a pair of words to the *S*, the shutter was lowered for a mean time of 2.034 sec. and was raised for a mean time of 1.950 sec. Between trials there was a short rest interval of approximately eight sec., the rest being measured from the time the last pair of words from the first trial disappeared until the stimulus word for the first pair in the second trial appeared.

The memory drum was mounted in a celotex screen which was designed to conceal the *E* from the *S* and, in general, to reduce the number of distracting stimuli which the *S* might receive.

On each trial, the score recorded was the number of correct anticipations, i.e., the num-

ber of response items given correctly before the shutter raised.

Subjects.—All of the *Ss* in the experiments were undergraduate students at the State University of Iowa who were required to participate in experimental work as a portion of their course in General Psychology. None of the *Ss* had previously served in a verbal learning experiment. No pre-training or trial runs were given to any *S* and no *S*, served in more than one condition of the experiments.[2]

Ss were discarded from the experiments if they failed to meet or exceeded the criteria for minimum and maximum acceptable performance noted below:

1. *Ss* were discarded if, by the end of the fifth trial of learning, they had failed to make a single correct response.

2. *Ss* were discarded if, by the end of the fifth trial, they had made a perfect score on one of the first five trials.[3]

Instructions.—All *Ss* received the State University of Iowa standard laboratory instructions for learning by the method of paired-associates.

TABLE I

THE CONDITIONS OF EXPERIMENT I

Condition	Amount of Original Learning	Length of Rest	Amount of Relearning	*N*
I	10 trials	no rest	10 trials	15
II	10 trials	1 min.	10 trials	15
III	10 trials	5 min.	10 trials	15
IV	10 trials	10 min.	10 trials	15
V	10 trials	20 min.	10 trials	15
VI	10 trials	40 min.	10 trials	15
VII	10 trials	24 hours[4]	10 trials	15

Experiment I.—The first experiment was designed to measure retention following various lengths of rest interval when original learning was for a constant number of trials. An attempt was also made to determine whether or not warming up effects occurred during relearning as a function of the length of the retention interval.

The conditions of Experiment I are outlined in Table I. Due to administrative difficulties involved in scheduling *Ss*, complete randomization of *Ss* with respect to conditions could not be accomplished. Twelve *Ss* were discarded from the experiment due to an inability to meet the criterion for minimum acceptable performance. There was no indication that a systematic difference existed between the conditions with respect to the number of *Ss* eliminated on this basis.

TABLE II

THE CONDITIONS OF EXPERIMENT II

Condition	Amount of Original Learning	Length of Rest	Amount of Color Naming	Amount of Relearning	*N*
A	10 trials	no rest	none	10 trials	15
B	10 trials	24 hours	none	10 trials	15
C	10 trials	24 hours	1 trial[5]	10 trials	15

Rest Interval Activity.—During the rest intervals, the *Ss* were required to read cartoons which had been clipped from various journals and mounted on 5 × 7 in. cards. This activity was performed with the *S* seated in front of the memory drum. Under the experimental conditions involving a 24-hour rest period, *Ss* were required to read cartoons for five min. following the completion of original learning and for five min. before the first trial of relearning.

Experiment II.—The conditions of Experiment II are outlined in Table II. Two of the conditions of this experiment were run as a portion of Experiment I. *Ss* were randomly

[2] This departure from traditional procedure was made because of the belief that amount of warming-up effect is a function of total amount of previous practice and number of previous resumptions of practice on similar tasks. For the same reason, it will be noted, original learning was carried out to a work rather than to a performance criterion.

[3] The second of these criteria proved to be of no practical importance, since no subjects were discarded under its provisions.

[4] Condition VII was scheduled with a two-hour leeway. The mean retention interval under this condition was 23 hours, 58 min.

[5] Conditions B and C were scheduled with a two-hour leeway. The mean retention interval under condition B was 23 hours, 58 min., while under condition C the mean retention interval was 23 hours, 42 min.

TABLE III

MEAN NUMBER OF CORRECT RESPONSES ON THE LAST ORIGINAL LEARNING TRIAL AND THE FIRST RELEARNING TRIAL IN EXPERIMENT I

Condition	Length of Rest	Mean Number of Correct Responses		Difference (M_{11}-M_{10})
		Trial 10	Trial 11	
I	no rest	6.80	7.53	+0.73
II	1 min.	6.80	7.13	+0.33
III	5 min.	6.93	6.93	0.00
IV	10 min.	6.67	6.53	−0.14
V	20 min.	6.80	6.13	−0.67
VI	40 min.	7.00	5.80	−1.20
VII	24 hours	6.27	4.40	−1.87

assigned as between conditions B and C, but condition A was conducted under conditions which prevented random assignment of *Ss* between it and conditions B and C. Six *Ss*, two under each condition, were eliminated from the experiment due to an inability to meet the criterion for minimum acceptable performance.

Since Experiment II was concerned with an attempt to decrease amount of forgetting by the technique of having the *Ss* engage in a warming-up activity prior to the recall of an originally learned activity, a description of the nature of this warming-up activity should be included. As has been previously noted, this activity should be highly similar to the learning task without providing specific practice on the originally learned activity. Furthermore, it is desirable that little learning of any kind should occur during the warming-up period in order that possible transfer of training effects may be avoided. The warming-up task which was selected for use was a color naming task, the colors being arranged in a list of 15 paired-associates and presented on the memory drum at the same rate of presentation as the materials to be learned. *Ss* were instructed not to attempt to learn the colors, but merely to name the second color of each pair of colors as it appeared in the window of the memory drum. Thus, a warming-up task was provided which was highly similar to the learning task but which, it was thought, did not provide practice which would increase the degree of learning on the originally learned activity.

RESULTS

Experiment I.—The means for the last trial of original learning (trial 10) and the first trial of relearning (trial 11) are presented in Table III. It will be seen that amount of retention is a negatively accelerated, decreasing function of the length of the retention interval. In order to determine whether or not significant differences in original learning existed among the seven experimental groups of this study, two analyses of variance were conducted, one for the number of correct responses on the last original learning trial and the other for the total number of correct anticipations during the 10 trials of original learning. In neither case did the obtained differences approach being statistically significant.

Since the various retention intervals were introduced between trials 10 and 11, the 11th trial scores were analyzed by the method of covariance, the total number of correct responses during original learning serving as initial measures. This analysis resulted in an F-ratio of 3.602 for 6 and 97 *df*, a value which is significant at the one percent level of confidence.

The warming-up effect has been defined in terms of the greater slope of the relearning curve as compared with the slope of the original learning curve at the same level of initial proficiency. Lindquist (5) has proposed a method for the analysis of the significance of such slope differences. If this test is applied to the extreme conditions of the present experiment (conditions I and VII), it should be possible to determine whether or not warming-up effects occur as a function of the length of the retention interval.[6]

[6] It may be wondered why this analysis was performed on only two of the seven conditions of Experiment I. However, no more conclusions would result from the seven condition than from the two-condition analysis. In either case, all that could be concluded as a result of a significant value of F would be that the amount of warming-up effect is some function of the length

Such an analysis was made of trials 11 to 20, inclusive, under conditions I and VII. The results ($F = 54.75$ for 9 and 252 *df*) clearly indicate the significance of the difference in the slopes of the relearning curves under these two conditions. However, this procedure favors the obtaining of a difference of this sort, since the slope of the acquisition curve changes with the level of proficiency (the acquisition curve was negatively accelerated). Therefore, it is necessary to compare the relearning trials under condition VII (24-hour rest) with that segment of the condition I (no rest) learning curve which starts at the same level of proficiency. Thus, a segment of the condition I learning curve must be selected which begins with a trial mean most closely approximating 4.40 correct anticipations (this being the mean number of correct anticipations for condition VII on the first relearning trial). Examination of the data reveals that trial 6 under condition I fulfills this condition most adequately. When the trial means for the 10 relearning trials under condition VII are compared with the means of trials 6 to 15, inclusive, of condition I, it is apparent that the performance under condition VII shows the greater rate of improvement. This comparison can be made from the data presented in Table IV. An analysis of the slope difference noted under the conditions of this comparison resulted in an *F*-ratio of 32.59 for 9 and 252 *df*, a value which is highly significant. It may be concluded then, that the warming-up effect occurs as some increasing function of the length of the retention interval.

Experiment II.—In Table IV are presented the mean number of correct anticipations for the 20 trials under the three conditions of Experiment II. In order to determine whether or not significant differences in original learning performance existed as between these three experimental groups, two analyses of variance, one for the scores on the last trial of original learning (trial 10) and the other for the total number of correct anticipations during the first 10 trials, were conducted. In neither case did the differences between the experimental groups approach being statistically significant.

Using the total number of correct anticipations during the first 10 learning trials as initial measures, three analyses of covariance were conducted with respect to scores on the 11th trial. The first of these compared conditions A (no rest) and B

TABLE IV

MEAN NUMBER OF CORRECT ANTICIPATIONS ON EACH TRIAL UNDER THE CONDITIONS OF EXPERIMENT II

Trial	Condition A (I of Exp. I)	Condition B (VII of Exp. I)	Condition C
1	0.00	0.00	0.00
2	0.53	0.53	0.20
3	1.67	1.07	1.00
4	1.73	2.07	1.27
5	3.07	2.73	2.33
6	3.80	3.13	2.87
7	5.60	4.07	3.73
8	5.27	3.93	4.67
9	6.07	5.19	5.00
10	6.80	6.27	5.80
	no rest	24-hr. rest, no warm-up	24-hr. rest, warm-up
11	7.53	4.40	6.07
12	8.53	5.86	7.67
13	8.87	7.53	8.40
14	10.07	8.60	8.27
15	9.73	8.73	9.47
16	10.47	9.60	10.00
17	10.83	9.53	10.13
18	10.67	10.47	10.67
19	11.13	10.33	11.07
20	11.33	11.33	10.93
$(M_{11}-M_{10})$	+0.73	−1.87	+0.27

of the retention interval. The two conditions for analysis must, of course, be selected independently of the experimental results.

(24-hour rest without pre-recall warming-up), and resulted in an *F*-ratio of 10.93 for 1 and 27 *df*, a value which is significant at the one percent level of confidence. A comparison between condition A (no rest) and condition C (24-hour rest with one warming-up trial before recall) yielded a non-significant *F*-ratio of 1.09. A comparison of condition B (24-hour rest without warming-up) with condition C (24-hour rest with one warming-up trial before recall) yielded an *F*-ratio of 7.93 for 1 and 27 *df*, a value which is significant at the one percent level of confidence. Thus, it may be concluded that a significant amount of forgetting occurred during the 24-hour rest period when the subject was not warmed-up before recall, and that the warming-up technique produced a significant reduction in the amount of this forgetting. On the other hand, the warmed-up group (condition C) did not differ significantly from the no-rest group (condition A). It should perhaps be noted that an estimate of the recovery from retention loss which is based entirely upon the relearning means would tend to be conservative, since some compensatory differences exist in original learning. If changes in performance between trials 10 and 11 are considered, this effect becomes more clearly marked. Tables III and IV show the amount of change in the mean number of correct anticipations between trials 10 and 11 for each of the conditions under Experiments I and II.

The results of the two experiments reported here indicate that amount of retention is dependent upon the degree to which an *S* is warmed-up at the time of measuring retention. In the ordinary retention and relearning situation, the *S* becomes warmed-up during the relearning trials, causing the relearning curve to have a steeper initial slope than would be predicted on the basis of proficiency level alone. The results of Experiment I indicate that amount of warming-up effect increases as some function of the length of the retention interval. The results of Experiment II indicate that if the *S* enters relearning after having been warmed-up on some similar task, retention will be better than if the subject had not received this warming-up.

A question naturally arises concerning the relative effectiveness of warming-up and habit formation during the early stages of relearning. The data in Table IV make such a comparison possible. Since the pre-recall warming-up trial given under condition C was of the same duration as each of the learning trials, the comparison referred to above can be made by comparing trials 11, 12, 13, etc., under condition C with trials 12, 13, 14, etc., under condition B. The amount of superiority of condition B under these conditions will reveal the role of additional habit formation in the initial trial of relearning, if it is assumed that the learning task and the warming-up task are equally effective in warming-up the *S*.[7] When this comparison is made, it will be noted that the two functions are practically congruent and that, therefore, the effect of one warming-up trial of color naming is practically equivalent to one additional trial, during relearning, of the originally learned material. The conclusion must follow, that habit formation during relearning is responsible for a relatively minor portion of the initial increase in performance proficiency.

[7] It would seem unreasonable to suppose, in any case, that the color-naming task could warm-up the subject for rote learning performance more adequately than could rote learning performance, itself.

Summary

The results of Experiments I and II may be summarized as follows:

1. In the verbal learning situation, the retention of a list of paired-associate adjectives was found to be a negatively accelerated decreasing function of the length of the retention interval.

2. The amount of warming-up effect (differential slope of the corresponding portions of the original learning and the relearning curves) was demonstrated to be some increasing function of the length of the interpolated rest period.

3. When a warming-up activity is introduced immediately before recall, a significant reduction in amount of forgetting is obtained.

(Manuscript received
November 10, 1948)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**, No. 1.
3. Irion, A. L. The relation of 'set' to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
4. Jost, A. Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Z. Psychol.*, 1897, **14**, 436–472.
5. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
6. Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 142.
7. Thorndike, E. L. *Educational psychology*. New York: Teachers College, Columbia Univ., 1914, Vol. 3.
8. Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 220.

REPETITION OF TWO BASIC EXPERIMENTS ON REMINISCENCE IN SERIAL VERBAL LEARNING[1]

BY CLAUDE E. BUXTON

Northwestern University

At the time these experiments were done it was the intention of the author to repeat the reference experiment of Ward (9), then to employ this experimental design in studies of reminiscence in materials other than nonsense syllables and in lists of other lengths. It was impossible to complete the systematic series of studies at that time, and since then the reports of Melton and Stone (7), and Wilson (10) have extended our information about reminiscence in both syllable and adjective learning. The present paper, therefore, serves only to place on record certain substantiating evidence, secured in two separate experiments employing identical designs. Theoretical discussion and interpretation will be minimized. Procedure will be described jointly for the two experiments, but the results will be reported separately.

Procedure

Subjects.—Six experimental sessions are required to duplicate the rest intervals employed by Ward, and a certain amount of practice must precede these sessions. Because of this large time requirement, it was impossible to use the same *S*s for both syllable and adjective experiments. Therefore, the less desirable plan of using two independent groups was followed. Each group contained 18 college students, both men and women; all were volunteers from elementary psychology classes, and all were new to psychological experimentation.

Lists.—Ten lists of syllables and 10 of adjectives, each containing 16 items, were prepared. All lists were suitable for serial anticipation learning, by pronouncing and by a non-correction procedure (as for Ward; see 1). Lists of unrelated adjectives were taken from Melton's materials (6), and were constructed according to rules such as those described by Melton and Stone (7, p. 296). Syllable lists were constructed according to the usual conventions concerning initial and final letters, spacing of repeated sounds or letters, etc. They were also taken from Melton's materials; their mean association value was 27 percent, with the range of means among lists being from 24 to 32 percent.

All lists were presented by means of 35-mm. film belts in a low-power SVE projector. Exposure of successive items was controlled by a synchronous motor and cam. Projection was from the rear of a 2 × 3 ft. translucent screen. *S* sat at a distance of about seven ft. from the screen, and was presented with words or syllables some three in. high and easily read without glare in normal daytime room illumination.

The exposure rate was 2 sec., with 4 sec. and a cue symbol, i.e., a total of 6 secs., between successive presentations of a list.

Rest-interval activity.—Near *S* was a collection of light reading materials (*New Yorker* magazine, etc.). Whether the rest-interval was short or long (*S* had been told that he could not and should not guess how long it would be), *S* read or scanned one or more magazines. He began this activity when *E* stopped the projector, and was recalled from it by a warning word from *E*. Initially, and during the practice days, he was instructed not to rehearse during any rest interval (11).

Experimental design.—The 10 lists to be used with either group of *S*s were used always in the same order for that group, i.e., variations due to lists and to practice effects were confounded. The first four lists were learned in successive practice sessions. Any list was first carried to the mastery criterion of 16/16; then a rest interval was given (varying in length but never the same as any experimental interval), and finally *S* was recalled to the list again. In this way *S* practiced shifting to and from the rest-interval activity.

The experimental variable, for either group independently, was introduced by varying the length of rest during the six experimental days. The intervals (conditions) were: 6 sec. (control—

[1] These data were collected in 1940–41 at Northwestern University, and were analyzed in part with the aid of research assistance at the University of Iowa during 1945.

no pause in the learning), 30 sec., 2 min., 5 min., 10 min. and 20 min. In all conditions the pre-rest criterion of mastery was 16/16, with the immediately succeeding trial, whether after six sec. or one of the longer intervals, serving as the recall trial. *S* always continued practice after rest until he had relearned to 16/16. Six randomized orders of conditions were established, and three *Ss* assigned to each order. These orders were chosen so that each condition fell equally often in each position in the series, and approximately equally often before and after every other condition.

Results, Experiment I, Syllables

Practice-effects.—That the four practice days probably brought the learning-how-to-learn curve down considerably is shown by difficulty of learning to the pre-rest criterion of 16/16 during successive experimental days. The mean numbers of trials to the criterion, together with their SD's, were: 21.1 (10.2), 16.3 (7.7), 15.1 (5.6), 15.7 (6.4), 16.9 (8.0), 13.1 (5.6). Although list difficulty is also a determiner of these values, inspection suggests that only the first and last experimental days showed much of a departure from a constant pre-rest learning rate. There are small shifts in the means from day to day, but the experimental conditions were distributed equally often to days, and it may be assumed that the only effect of inequalities is to lower the precision of comparisons, not to bias them.

Recall data.—In accordance with an earlier definition (**2**), it is assumed here that recall data afford the most legitimate picture of reminiscence phenomena. In Table I are recorded the recall scores after varying lengths of rest. The data are graphed to show changes from performance in the control condition, in the upper part of Fig. 1. Inspection of table and graph indicates that as for Ward (**9**, p. 17), with recall in the control condition as a reference point, the retention curve for the total list rises for the 30-sec. and 2-min. rest intervals, has dropped almost to the level of the control condition at five min., and continues to drop for the remaining lengths of interval.

As a function of both the experimental design and the internal characteristics of the data, statistical analyses of the just-mentioned results are limited. (a) The variances about experimental means within the syllable data are heterogeneous, as shown by Bartlett's test, at the .01 level of confidence (cf. Snedecor, **8**, p. 249). This is in some measure a matter of experimental design for, in using the pre-rest criterion of one perfect recitation, a situation is created where variability (and degree of skewness of distributions) is a function partly of how far below the 'ceiling' of 16 the recall scores tend to drop. Ward used this same design, with a criterion of 12/12 syllables, but in his case retention apparently dropped enough that the heterogeneity-of-variance problem did not occur. (His SE's are almost identical for mean recalls after all experimental rest intervals.)

TABLE I

Recall Trial Means (with SD's) for Syllables, After Varying Lengths of Rest Period (N = 18 for Each Mean)

	6 sec.	30 sec.	2 min.	5 min.	10 min.	20 min.
Mean	13.72	14.89	15.39	13.83	13.22	10.94
SD	2.22	.79	.73	1.00	1.00	2.06

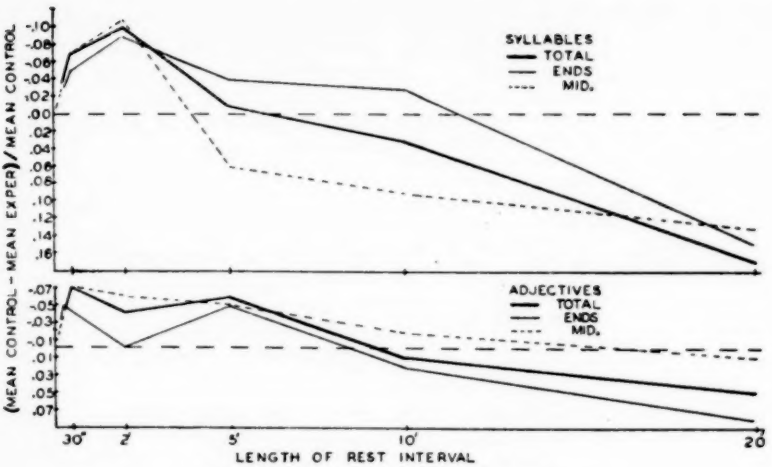


FIG. 1. Relationship between amount retained (relative to control condition) and length of rest interval

(b) A further consideration concerning statistical analyses is that the retention curve is complex, being partly in the reminiscence direction (shorter rest intervals) and partly in the forgetting direction (longer intervals). Therefore, analyses of variance including all points are not very informative.

It is suggested by Cochran (4, p. 28) that when variances are heterogeneous, the "validity of the *F*-test for all treatments is least affected." Following this suggestion, analyses of variance for the control condition and for (only) the points lying above it have been performed. The error estimate was a residual from which had been removed the variance due to individual differences. For scores based on the total list of syllables, the apparent reminiscence is probably statistically significant ($F = 7.00$, 3/51 *df*; .01 level = 4.19). With the caution made necessary by the known characteristics of the distributions in question, it seems reasonable to conclude that reminiscence did appear in the syllable lists.

Serial position phenomena in recall. —Since the relationship between performance and serial position is an important aspect of current theorizing about reminiscence, one characteristic of serial position data from the present experiment will be reported. Two

TABLE II

RECALL TRIAL MEANS (WITH SD'S) FOR SYLLABLES, FOR END POSITIONS AND MIDDLE POSITIONS, AFTER VARYING LENGTHS OF REST PERIOD ($N = 18$ FOR EACH MEAN)

	6 sec.	30 sec.	2 min.	5 min.	10 min.	20 min.
Mean, end	5.39	5.67	5.94	5.61	5.56	4.50
SD	.82	.54	.33	.69	.53	1.25
Mean, middle	4.94	5.33	5.61	4.61	4.39	3.17
SD	1.41	.77	.60	.90	.93	1.15

TABLE III

RECALL TRIAL MEANS (WITH SD'S) FOR ADJECTIVES, AFTER VARYING LENGTHS OF REST PERIOD (N = 18 FOR EACH MEAN)

	6 sec.	30 sec.	2 min.	5 min.	10 min.	20 min.
Mean	13.61	14.67	14.22	14.50	13.44	12.78
SD	2.24	1.57	1.53	1.46	2.06	2.62

scores, one for end syllables (positions 1–2–3–14–15–16) and one for middle syllables (positions 6–7–8–9–10–11), were assigned to each recall performance by each *S*. Then these scores were tabulated according to length of rest interval before the recall. Table II shows the empirical data, and the upper part of Fig. 1 pictures the derived percent changes in successive recall means for the two groups of items.

In every instance (see Table II), end syllables were retained better than middle items, a fact consistent with the usual expectations based on initial level of mastery. However (see Fig. 1), when the mean recalls after rest are considered in relation to their reference point, the control mean, the initial rise in middle positions seems greater than in end positions. Tests of significance are therefore desirable. There is some heterogeneity of variance, although in Table II this is less apparent than when total scores were used. Analyses of variance were therefore made, again using as an error estimate a variance from which individual differences had been removed. These analyses were performed for means of end and middle positions separately, to avoid complications by such heterogeneity of variance as appeared to exist. Briefly, neither the end or middle positions taken separately show a significant tendency toward reminiscence (although the F-value for end positions is the larger of the two, possibly because the variances about mean recalls are smaller than for middle position items). Such significant reminiscence as appeared in total syllable lists (see Table I) evidently depended upon both end and middle serial positions, for either taken by itself did not exhibit reminiscence significantly.

RESULTS, EXPERIMENT II, ADJECTIVES

In general, the analyses to be reported were the same as for Experiment I, and descriptions of statistical techniques, etc., will not be repeated.

Practice-effects.—During the experimental days, the numbers of trials to the pre-rest criterion of 16/16 for adjective lists were, together with their SD's: 13.8 (3.9), 10.6 (4.4), 12.6 (5.3), 11.4 (3.9), 11.7 (3.7), 9.7 (3.4). In Experiment II, then, the learning-how-to-learn curve is approximately level during the experimental sessions. It will be noted that these values are definitely smaller than the comparable ones in Experiment I.

Recall data.—The recall data for adjectives are shown in Table III and in the lower part of Fig. 1 (marked total list). In these lists, retention improves for the three shorter rest intervals but not, in the present data, according to a smooth function; in addition, the recall scores do not drop much below those for the control condition, even at the 20-min. rest length. The adjective retention

curve thus seemingly agrees only in part with that of Melton and Stone: they found only indications of positive acceleration in forgetting, as rest interval length increased, whereas the present curve rises somewhat at the shorter intervals and, by inference, shows next the positive acceleration phase to which Melton and Stone refer. That is, the shape of the initial portion of any forgetting curve is a function of both the forgetting process and any opposed process. Seemingly, if the opposed process is very potent, the curve will rise or be maintained level for a while, and then the forgetting curve must be positively accelerated for a while. Because of the high level of initial mastery required in the present study, one might expect the negative acceleration phase to appear with a shorter rest interval than for Melton and Stone. Such was the case, for negative acceleration is probably present before 20 min. (see lower part, Fig. 1) in the present study, whereas positive acceleration seemed to be present throughout the Melton and Stone data.

Since the retention curve for adjectives seems to rise initially, in the present data, it is of particular interest to know whether this rise is a significant one statistically. Within the limitations imposed by heterogeneity of variance, it is possible to conclude that the results are actually not opposed to those obtained by Melton and Stone, for the apparent reminiscence was not statistically significant ($F = 1.60$, $3/51$ *df*; .05 level = 2.79). Since the forgetting reported by Melton and Stone for their shorter rest intervals was also not significant, it can only be inferred at the present time that the shorter rest intervals lead to insignificant changes in recall of adjective lists. (See Wilson, 10, also.)

Serial position phenomena in recall. —Table IV shows mean recall scores for middle and end adjectives (where middle and end positions are defined as in Experiment I), and percent changes in these means are graphed in the lower part of Fig. 1. Although Fig. 1 suggests that the middle items in a list tend to show more reminiscence than the end items, neither position produced statistically significant evidence of reminiscence.

Discussion

In the preceding separate summaries of results from the two experiments it was not thought necessary to cite comparative data on learning or re-learning. It will suffice here to remark that (a) the difficulty of learning to the pre-rest criterion was well equated among the conditions within each experiment; (b) syllables were, by inspection, more difficult to learn than adjectives; and (c) percent savings (based on trials to the criterion, and to one perfect trial after rest) did not show any consistent

TABLE IV

Recall Trial Means (with SD's) for Adjectives, for End Positions and Middle Positions, After Varying Lengths of Rest Period ($N = 18$ for Each Mean)

	6 sec.	30 sec.	2 min.	5 min.	10 min.	20 min.
Mean, end	5.61	5.94	5.61	5.88	5.49	5.11
SD	.66	.26	.66	.53	.81	1.14
Mean, middle	4.77	5.16	5.11	5.05	4.88	4.72
SD	1.41	1.03	1.04	.93	1.31	1.36

pattern of benefit from the interpolated rest. This resulted partly from the high pre-rest criterion of mastery, which narrowed the range of re-learning trials within which the rest might show its influence.

In general, it will be seen that the main results of both Ward and of Melton and Stone are substantiated in the total-list recall data of the two experiments reported here. The fact that reminiscence appears with syllables but not with unrelated adjectives is suggestive in itself, but it must be checked in further experiments where a lower criterion of mastery is employed and the same *Ss* learn both types of materials. The present data do not show the serial position phenomena often thought to be associated with the appearance of reminiscence, nor do they show the re-learning benefits found in some earlier work.

From the methodological point of view, the present data demonstrate convincingly the desirability of using a pre-rest mastery criterion which is well below the 'ceiling' or perfect score.

Summary

1. Two complete Ward-type reminiscence experiments were performed, with independent groups of *Ss*. One group learned unrelated adjectives, the other nonsense syllables. Each group, after first reaching the criterion of 16/16, had rest intervals on various experimental days, of 6 sec., 30 sec., 2, 5, 10, or 20 min., in a randomized order for each *S*.

2. The syllable recall data showed an initial significant rise, then a fairly sharp fall, as length of rest interval increased. This substantiates Ward's principal finding.

3. Recall of adjectives showed an insignificant initial improvement, then a small decrement, as length of rest interval increased. Essentially, the data are in agreement with Melton and Stone, and Wilson, in showing no significant evidence of reminiscence with unrelated adjectives, at least under the procedures used to date.

4. Middle positions, in the syllable lists, upon inspection, seemed to show more reminiscence than end positions. Tests of significance showed, however, that neither position by itself showed significant reminiscence.

5. A similar result was found in serial position analyses of the adjective data.

6. Data on original learning showed the to-be-expected difference in difficulty of adjective and syllable lists; a single interpolated rest did not lead to re-learning benefits which varied as a function of rest. (These data not reported in detail.)

(Manuscript received
November 30, 1948)

References

1. Buxton, C. E., & Bakan, Mildred B. Correction vs. non-correction techniques as related to reminiscence in serial anticipation learning. *J. exp. Psychology*, 1949, **39**, 338–341.
2. Buxton, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, **40**, 313–340.
3. Buxton, C. E., & Ross, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
4. Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.
5. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
6. Melton, A. W. Materials for use in experimental studies of the learning and reten-

tion of verbal habits. (Unpublished manuscript, 1940)
7. MELTON, A. W., & STONE, G. R. The retention of serial lists of adjectives over short time intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
8. SNEDECOR, G. W. *Statistical methods.* (4th ed.) Ames: Iowa State College Press, 1946.
9. WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 220.
10. WILSON, J. T. The formation and retention of remote associations in rote learning. Ph.D. thesis, Stanford Univ., 1948.
11. WITHEY, S., BUXTON, C. E., & ELKIN, A. Control of rest interval activities in experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 173–176.

ATTEMPTED PUPILLARY CONDITIONING AT FOUR STIMULUS INTERVALS[1]

BY ERNEST R. HILGARD, CHARLES E. DUTTON, AND JOHN S. HELMICK[2]

Department of Psychology, Stanford University

Introduction

The significance of pupillary conditioning for learning is enhanced because of the conjectures that responses under the control of the autonomic system may be learned according to a different paradigm from that appropriate to responses of striate muscle (e.g., **9**, **10**, **14**). The opposing conjecture is that all conditioning is of one kind (**8**).

Knowledge of the relationships within pupillary conditioning is in a most unsatisfactory state. Although positive results were reported in some studies (e.g., **1**, **2**, **7**), the negative results obtained by others are disturbing (e.g., **6**, **11**, **12**, **13**, **15**). The positive results with animals have been with what may be called the emotionally-activated pupil, in which electric shock (not visual stimulation) served as the unconditioned stimulus (e.g., **3**, **4**, **5**). At least one recent attempt to condition the pupillary responses of animals (rabbits), with a visual stimulus as the unconditioned agent, has yielded negative results followed by negative results with human subjects under 10 conditions of experimentation (**13**).

It has seemed desirable, therefore, to undertake an orderly exploration of the variables conventionally supposed to modify the conditioning process in order to determine the circumstances under which pupillary conditioning is favored. In the present study the variable introduced is that of the time interval between conditioned and unconditioned stimuli. Because much autonomic conditioning has made use of intervals in seconds rather than fractions of seconds, this first exploration was of intervals of 0, 3, 6 and 9 sec. between the onset of the conditioned stimulus and the onset of the unconditioned stimulus. Because the conditioned stimulus endured until the unconditioned stimulus was terminated, the arrangement at the longer intervals is technically that of delayed conditioning rather than of trace conditioning.

Although the results are chiefly negative, they are not without interest on that account. The experimental arrangements are such that pupillary conditioning ought to be obtained readily according to theories which base conditioning upon the association between conditioned stimuli and concomitant unconditioned reflexes. If satisfactory conditioning cannot be established it means that the classical theory of conditioning will be in need of revision. This matter is of sufficient theoretical importance to warrant a continued search for the circumstances favorable to pupillary conditioning.

Experimental Arrangements

The apparatus is essentially that described by Hilgard, Miller and Ohlson (**6**). The pupil was viewed through a telescope equipped with parallel vertical hairs manually adjusted to provide a registration of pupil diameter on a wax-paper polygraph. The records yielded a magnification 10 times the actual pupillary diameter.

[1] This investigation was supported in part by a grant from the research fund of the School of Social Sciences, Stanford University.

[2] Now at the University of California, Los Angeles, California.

The unconditioned stimulus (UncS) was an increase in illumination on a flash glass surface 10 cm. × 12 cm. located at a distance of 15 cm. from the eyes, just below the line of vision. The resting illumination was 60 ml. An increase to 1000 ml. served as the unconditioned stimulus. Binocular fixation was maintained upon an illuminated cross which was seen in a mirror so as to appear one meter directly ahead of the subject, and just above the illuminated patch. The change in illumination was an adequate unconditioned stimulus for all subjects, producing a prompt pupillary constriction averaging 1.5 mm., with a range from .9 to 2.8 mm., and a standard deviation .5 mm. for the 43 subjects participating in the experiment.

The conditioned stimulus (CS) was the sounding of an ordinary high frequency buzzer, well above threshold.

The presentation of stimuli was automatically controlled. One cycle of the interval timer required 35 sec., so that, unless the cycle was interrupted, a trial occurred every 35 sec. The experimenter occasionally interrupted the cycle of the interval timer to introduce rest periods, and he controlled the switches determining which stimuli would appear. When they did appear, and their duration, were controlled by the automatic features of the interval timer. Therefore, the primary attention of the experimenter was devoted to the following of the pupillary diameter.

Subjects were students, both men and women, from the introductory psychology course. Each subject served for two days. Trials within the two days were arranged as follows:

Day 1
6 trials with CS alone
40 reinforced trials (CS + UncS), with 8 interspersed test trials (CS alone).

These trials were presented in banks of 6 trials with rests between each bank. One test trial occurred in each bank.

Day 2
40 reinforced trials, with 8 interspersed test trials, as on Day 1
6 trials with CS alone

There were four groups of subjects, with conditions alike except for the interval between stimuli. The four groups were as follows:

Stimulus Interval	No. of Subjects	Description
0 sec.	11	Sound (CS) and Light (UncS) begin together, and both continue for 3 sec.
3 sec.	11	Sound (CS) begins 3 sec. before Light (UncS), and both continue together for 3 sec.
6 sec.	11	Sound (CS) begins 6 sec. before Light (UncS) and both continue together for 3 sec.
9 sec.	10	Sound (CS) begins 9 sec. before Light (UncS) and both continue together for 3 sec.

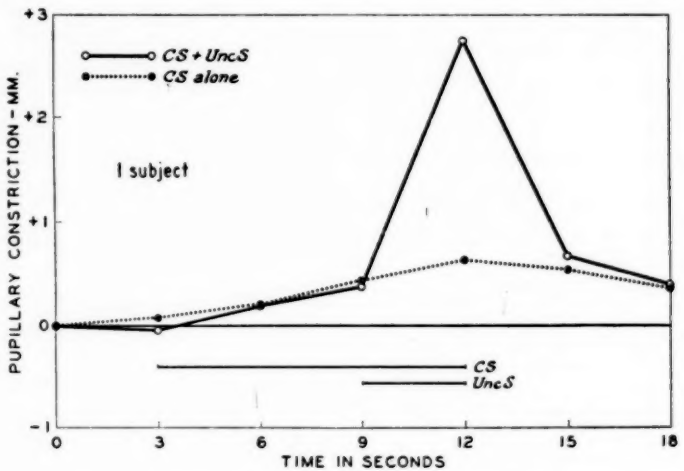


FIG. 1. Pupillary conditioning within reinforcement trials, Subject 28. The onset and duration of both the buzzer (CS) and the light (UncS) are shown at the bottom of the figure. The solid line gives the mean values for responses on reinforced trials, the dotted line for interspersed test trials.

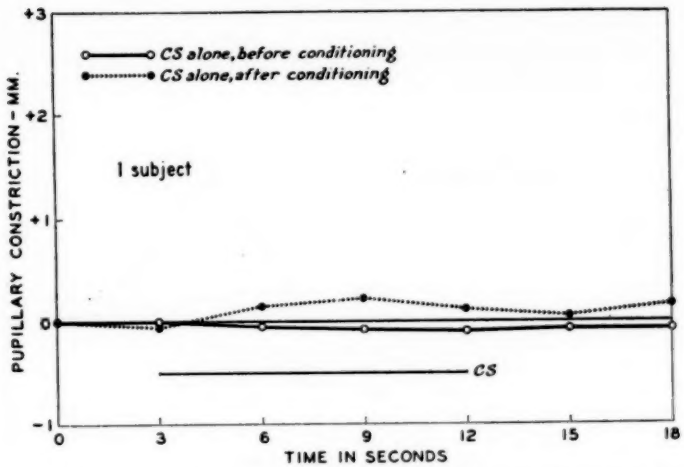


FIG. 2. Pupillary responses before and after conditioning, Subject 28. The solid line represents pupillary changes to the buzzer (CS) within six trials before any reinforcement, the dotted line responses within six trials at the end of two days' conditioning.

While continuous records were taken within each bank of trials, a standard practice was adopted in quantifying the records. The beginning of the trial was selected arbitrarily as nine sec. before the onset of light. The pupillary diameter at this moment was taken as a base-point and all later deviations were measured from it. Measurements were taken at successive three-second intervals, regardless of the configuration of responses between the three-second points. This was found desirable because of the number of fluctuations of the pupil unrelated to the stimuli. By taking measurements from a zero point which had nothing to do with pupillary responses, such 'chance' fluctuations should be balanced out, though systematic pupillary drifts will of course be registered. It is inevitable that the resting diameter of the pupil will change because of adaptation throughout an experiment in which intervals of relative darkness are followed by occasional periods of high illumination. By making all measurements relative to the size near to the onset of the stimuli for a given trial, the gross changes in resting diameter are disregarded. The records show actual diameter, so that any artifacts introduced by the arbitrary method of measurement can be studied. Observation of the records showed that results would not have differed had other points been selected for measurement.

The apparatus was moved to a new laboratory in the midst of the experiment, and slight technical improvements were made. Some of the subjects of every group were experimented upon in the new laboratory, and there is no evidence to indicate any change in results as a consequence of moving.

RESULTS FOR A SUBJECT SHOWING CONDITIONED RESPONSES

Of the 43 subjects completing the experiment only one (Subject 28) yielded results which from a statistical standpoint gave satisfactory evidence of conditioning. This subject is presented in defense of the contention that the recording method was appropriate to the detection of conditioning, had it occurred.

The main results for this subject are presented in Figs. 1 and 2.

Fig. 1 shows the course of pupillary changes within the reinforced trials, as measured both from the reinforced trials themselves and the non-reinforced trials interspersed among them. The consistency of the results on the two sets of trials gives an indication of the stability of the data for the mean responses of single subjects.

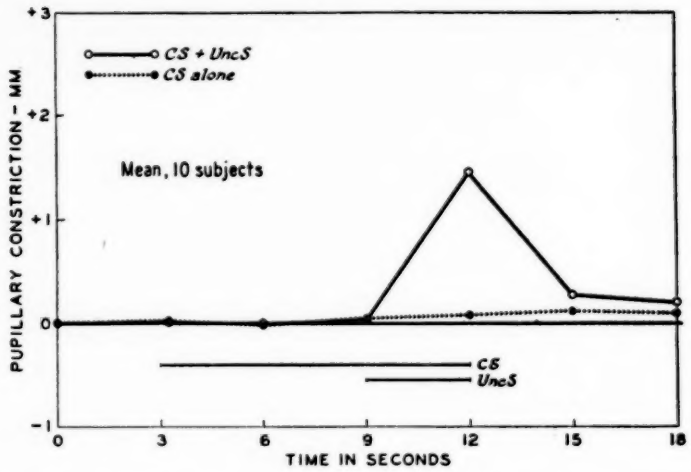


FIG. 3. Pupillary conditioning within reinforcement trials, 10 subjects. Mean responses of all subjects of the 6-sec. group, excluding the subject whose responses are plotted in Fig. 1. See caption of Fig. 1.

For this subject the pupil begins to constrict after the buzzer (CS) begins to sound, before the light (UncS) comes on, thus showing the familiar anticipatory character of conditioned responses when stimulus intervals permit. The conditioned constriction of 0.4 mm. by the time of onset of the light is $\frac{1}{7}$ of the total constriction of 2.8 mm. to the unconditioned stimulus. Even such a moderate amount of conditioning is adequately registered.

Fig. 2 contrasts the changes in response to CS before and after the reinforcements have been given. The departure of the two curves is evidence for conditioning.

When the 22 individual values determining the 12-sec. points in the dotted lines of Figs. 1 and 2 are studied statistically, the mean is found to be farther above the baseline than would be predicted by chance. The *t*-value is 5.9, well beyond the .01 level of significance, which for 21 degrees of freedom requires a *t* of 2.8. Significant pupillary constriction at the 12-sec. point, as here demonstrated, is the indicator of conditioning which is being used in the remainder of this paper.

Even the results for this one subject are none too successful as a demonstration of pupillary conditioning. The conditioning is very prompt, and not progressive. There is somewhat greater responsiveness within the first half of each day. It is possible that the trials were somewhat crowded to be most advantageous. Within the six trials to the conditioned stimulus alone at the end of conditioning there was no consistent trend toward extinction. The conditioned response of greatest magnitude was found on the fourth of these trials. These considerations are not crucial, and would have value regarding the nature of pupillary conditioning if generalizations could be made from the results of more subjects. Unfortunately, this subject is the only one for whom any positive statements whatever can be

made about the course of conditioning or extinction.

One possible reason for the uniqueness of this subject is worth noting. Her pupillary constriction was the greatest of any subject tested, averaging 2.8 mm. as compared with the mean of 1.5 mm. for all subjects. This may suggest that stronger unconditioned stimuli should be used in further attempts at conditioning, although some other factor in individual responsiveness may be involved. There were four other subjects with responses averaging above two mm. None of them showed any conditioning.

Evidence Against Pupillary Conditioning at the Stimulus Intervals Used

The one subject reported illustrates the kinds of data available for every subject. In the further analysis, each subject contributes equally by having his mean values for each point enter as a single value in the group mean. In the further computations, the *N* used is the number of subjects.

The negative results for the groups of subjects can be illustrated by the data for the others of the 6-sec. group to which Subject 28 belonged. These data are plotted in Figs. 3 and 4, with the same scale and the same measures as plotted in Figs. 1 and 2. For this purpose Subject 28 has been excluded, although in further statistical treatment she is included as a regular member of the group.

The consistency with which the points fall on the baseline in both Figs. 3 and 4 shows both the lack of conditioning and the adequacy of the method of treatment in cancelling out meaningless fluctuations in pupillary diameter. There is slight drift toward constriction, which agrees with the more exaggerated drift for the single subject of Fig. 1.

Because of the consistency of no response, statistical demonstration is almost unnecessary. Yet the precision of measurement is such that very little response would be necessary

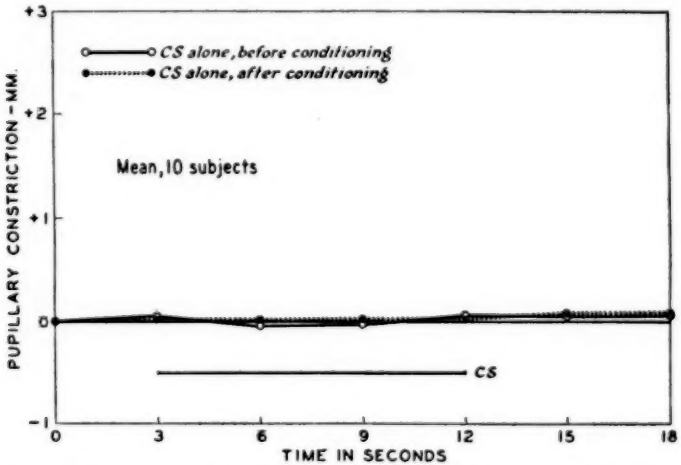


Fig. 4. Pupillary responses before and after conditioning, 10 subjects. Mean responses of all subjects of the 6-sec. group, excluding the subject whose responses are plotted in Fig. 2. See caption of Fig. 2.

for significance, so that the usual statistical tests are worth applying. For the purposes of statistical appraisal we have selected the mean change in pupillary size from the arbitrary zero sec. (preceding the onset of the UncS by 9 sec.) to the time of termination of both CS and UncS, 12 sec. later. These measures were yielded by the 16 test trials during reinforcement, and the 6 final test trials, a total of 22 trials for each subject. Every effort was made to select a point favorable for the detection of conditioning, had conditioning been evidenced. The results for the four groups are given in Table I.

What we find from Table I is that the pupillary constriction which might indicate conditioning reaches the .05 significance level at the 3-sec. and 6-sec. intervals, but fails in either case to reach the .02 level. Had a correction been made for the slight drift toward constriction from the beginning to the end of the records, the results would have been even less favorable to the indication of any conditioning at all.

The fact that the *t*-values are as close to significance as they are shows that the method would have detected responses of small magnitude had they been at all consistent. The presence of the one subject in the 6-sec. group with evidence for conditioning brought that group near to significance.

Pupillary conditioning might have produced differences in variability between groups. When tested by the *F*-ratio, we find for the groups yielding the largest difference, the 6-sec. and 9-sec. groups, the ratio $\sigma_6^2/\sigma_9^2 = 4.30$. This value exceeds the .05 level of significance, which for 10, 9 *df* is 3.13, but it does not reach the .01 value of 5.26.

Discussion

The results show that satisfactory conditioning was not obtained under arrangements which should have been conducive to it. Because one subject showed conditioning, and the test of significance was marginal for some of the groups, the possibility of pupillary conditioning cannot be ruled out. A search for favorable conditions must go on.

It should be pointed out that satisfactory conditioning requires more than the meeting of a test of statistical significance. Conditioning would be useless as an aid in understanding learning unless there were enough of

TABLE I

Significance of Possible Conditioned Responses Present on Test Trials Within and Following Two Days of Conditioning

	Interval Between Onset of CS and UncS			
	0 sec.	3 sec.	6 sec.	9 sec.
Number of subjects	11	11	11	10
Mean pupillary constriction, mm.	−.10*	.11	.13	.02
σ_M pupillary constriction, mm.	.053	.048	.054	.032
t	1.9	2.3	2.4	.6
t for significance at .01	3.17	3.17	3.17	3.25
t for significance at .02	2.76	2.76	2.76	2.82
t for significance at .05	2.23	2.23	2.23	2.26

*Minus sign indicates pupillary dilation.

it for differential effects to occur, as in generalization and discrimination. This calls for easily identifiable conditioned responses, with describable characteristics. Responses which meet the bare statistical test of 'probably present beyond chance expectation' hardly suffice. Even such responses were not found among these subjects as a group under the circumstances of these experiments.

Unless satisfactory pupillary conditioning can be obtained, classical conditioning theory will have to be revised. It is important that the negative results should not be allowed to stand until every effort has been made to discover more favorable conditions.

Summary

1. Pupillary conditioning was attempted on two successive days with each of 43 adult human subjects. The conditioned stimulus was a suprathreshold buzzer, the unconditioned stimulus an illumination increase.

2. The stimulus-intervals used with different groups included the buzzer and light presented together, and the buzzer preceding the light by 3, 6, and 9 sec.

3. Successful conditioning with one subject at the 6-sec. interval demonstrated that the recording method was satisfactory to detect conditioned responses of small magnitude.

4. Conditioning was generally unsatisfactory at all four stimulus intervals. Although at the 3- and 6-sec. intervals, responses reached the .05 level of significance, no group reached the .02 level.

5. It is pointed out that bare statistical significance (had it been reached) would not have sufficed to make pupillary conditioning useful in the study of learning. The search for more satisfactory circumstances for pupillary conditioning is recommended. Continued negative results would require a revision in classical conditioning theory.

(Manuscript received
December 7, 1948)

References

1. Baker, L. E. The pupillary response conditioned to subliminal stimuli. *Psychol. Monogr.*, 1938, 50, No. 223.
2. Cason, H. The conditioned pupillary reaction. *J. exp. Psychol.*, 1922, 5, 108–146.
3. Girden, E. The dissociation of pupillary conditioned reflexes under erythroidine and curare. *J. exp. Psychol.*, 1942, 31, 322–332.
4. Harlow, H. F. The effect of incomplete curare paralysis upon the formation of conditioned responses in cats. *J. genet. Psychol.*, 1940, 56, 273–282.
5. Harlow, H. F., & Stagner, R. Effect of complete striate muscle paralysis upon the learning process. *J. exp. Psychol.*, 1933, 16, 283–294.
6. Hilgard, E. R., Miller, J., & Ohlson, J. A. Three attempts to secure pupillary conditioning to auditory stimuli near the absolute threshold. *J. exp. Psychol.*, 1941, 29, 89–103.
7. Hudgins, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. gen. Psychol.*, 1933, 8, 3–51.
8. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
9. Mowrer, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harvard Educ. Rev.*, 1947, 17, 102–148.
10. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
11. Steckle, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, 15, 369–377.
12. Steckle, L. C., & Renshaw, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, 11, 3–23.
13. Stern, F. *An investigation of pupillary conditioning.* Ph.D. dissertation, Univ. Washington, 1948.
14. Tuttle, H. S. Two kinds of learning. *J. Psychol.*, 1946, 22, 267–277.
15. Wedell, C. H., Taylor, F. V., & Skolnick, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, 27, 517–531.

'SPREAD' IN RECALLING FAILURE AND SUCCESS

BY JOSEF NUTTIN

University of Louvain (Belgium)

In the last years, several hypotheses have been set up to explain the so-called 'Thorndike-effect,' i.e., the fact that punished responses in the neighborhood of a rewarded one are more frequently repeated in the course of the learning process than the more removed responses are (**2, 3, 7, 10, 12, 15, 17, 19, 20**). In order to obtain further information about this puzzling phenomenon, the present author studied the problem from another point of view. The experiments to be reported here deal with the subject's *memories* of reward and punishment. The purpose is to examine whether 'punished' responses, immediately preceding or following a successful one, are more frequently *recalled* as 'rewarded' than more remote items are.

The results of previous experiments on recall in serial learning in general —carried out at the same laboratory several years ago—suggested the assumption that a punished response, situated *between two successful ones*, tends to be recalled more frequently as rewarded than a punished response situated *between two other punished connections* (**1**).

Therefore, special experiments were set up to examine whether a phenomenon of spread, analogous to the 'spread of effect,' can be said to exist among the *recalls* of failure and success in serial learning.

EXPERIMENT I [1]

The *S* is presented with a printed list of 40 well-known stimulus words. Each stimulus word is accompanied (at its right) by five two-syllable words in an 'unknown language' (actually, these words are artificially constructed). For example:

HAND—	kos.var	em.sos	den.sok	ger.sem
PEN —	nos.ga	tinlso	ves.kor	hu.ris

The *S* is told that one of the five strange words is the exact translation of the stimulus word, and he is asked to learn the translation words in the following way. After reading the first stimulus word, he pronounces one of the four nonsense words at his choice. The *E* tells him if his choice is 'right' or 'wrong,' *according to a prearranged pattern of success and failure distribution.* After this, the *S* goes on immediately to the second stimulus word, pronounces one of the four following nonsense words and hears the *E*'s 'right' or 'wrong,' etc. (at the rate of five sec. per word).

The *S* is told that the whole series is to be repeated in the same way, in order to learn perfectly the correct response for each stimulus word. However, after finishing the first presentation of the whole series, the instructions are changed. The *S* is asked now to *reproduce* or to repeat for each stimulus word read again, the *same* translation word he has just given during the first presentation of the series, *adding whether his choice was said to be right or wrong* (memory of punishment and reward).

The prearranged pattern for calling responses 'right' or 'wrong' provided a certain number of punished connections at different distances (steps) from a rewarded response. Three different patterns were set up (A, B, and C), providing different schemes in distributing reward and punishment. The number of punished items at the different 'steps' is indicated in Table I for the Patterns A, B, and C. Experiment I was carried out with 12 groups of male *Ss* differing in age (12–20 years) and occupation (total number of *Ss*: 358).

Results

The strength of the connections between the stimulus and the nonsense word, as measured by the percentage of exact reproductions, does not concern us at the moment. We wish only to examine the percentage of recalls of failure and success, with

[1] The experiments described in this paper were performed, in part, with the assistance of V. Van Cutsem, Paed. D., and O. Degadt, Paed. Lic.

TABLE I

NUMBER OF RESPONSES PUNISHED AT THE DIFFERENT DISTANCES FROM REWARD IN EXP. I AND EXP. II (*N* PER *S*).

	Experiment I Pattern			Exp. II
	A	B	C	
Between two rewards	8	1	1	6
One step after/before reward	9	13	13	9
Two steps removed from reward	7	8	10	8
Three steps removed	2	5	5	6
Four steps removed	0	3	2	5
Five or six steps removed	0	2	1	4
For comparison:				
Responses rewarded	14	8	8	12
Total	40	40	40	50

respect to the proximity of the response to a rewarded connection.

Table II indicates the percentages of responses, at the different degrees of remoteness from reward, that the *S*s recalled as having been 'rewarded' during the first presentation of the series.

Comparing the percentages of the more extremely situated items with the closer ones ('one step' cases vs. 'three steps and more'), we obtain the results as indicated in Table III.

The total results of the 358 *S*s show that 27.0 percent of the punished responses immediately following and preceding a reward are recalled as rewarded. This happens only in 22.1 percent of the cases that are three or more steps removed from reward. This difference is statistically significant.

No systematic difference was found between the items *preceding* and those *following* reward on the different steps. Therefore no separate men-

TABLE II

WRONG RESPONSES RECALLED AS REWARDED, AT DIFFERENT DISTANCES FROM REWARD AS STATED

Groups	*N*	*N* of Respons. Involved	Wrong Responses					
			Between 2 Rew. %	1 step aft./bef. %	2 steps removed %	3 steps removed %	4 steps removed %	5 steps removed %
A_1	35	1400	39.3	35.5	33.9	31.4	—	—
A_2	37	1480	40.5	32.1	29.3	36.5	—	—
B_1	13	520	30.8	14.8	23.9	13.9	12.8	15.4
B_2	14	560	57.1	23.6	26.8	24.3	16.7	7.1
B_3	19	760	31.6	22.7	23.7	24.2	19.3	10.5
B_4	40	1600	15.0	22.5	19.4	20.5	15.0	12.5
C_1	40	1600	27.5	17.6	24.2	20.6	19.0	22.5
C_2	40	1600	30.0	25.9	24.5	25.6	22.0	22.5
C_3	40	1600	45.0	24.8	25.0	15.6	21.0	20.0
C_4	40	1600	25.0	20.4	15.5	20.0	16.0	20.0
C_5	40	1600	30.0	29.1	20.0	23.9	17.0	25.0

TABLE III

RESPONSES RECALLED AS REWARDED AMONG THE 'ONE STEP' CASES VS. THE 'THREE STEPS AND MORE' CASES

(Percentages of right responses recalled as rewarded are added for comparison)

Groups	Right Resp. (N = 3296) %	Wrong Responses		Diff. between 1 step and 3 or more
		1 step removed (betw./aft./bef.) (N = 5228) %	3 steps and more removed (N = 3108) %	
A_1	46.7	38.0	33.3*	+4.7
A_2	56.9	36.1	30.9*	+5.2
B_1	20.2	15.9	13.7	+2.2
B_2	27.7	26.0	18.6	+7.4
B_3	21.1	23.3	20.0	+3.3
B_4	32.2	22.0	17.2	+4.8
C_1	28.5	18.3	20.3	−2.0
C_2	34.7	26.2	24.1	+2.1
C_3	29.7	26.2	17.8	+8.4
C_4	25.9	20.7	18.7	+2.0
C_5	31.9	30.9	21.9	+9.0
Mean (weighted)		27.0 SD = 0.7% SD (diff.) = 1%	22.1 SD = 0.8% (CR = 4.9)	

* This percentage includes the 'two steps' cases, because only two items in Group A_1 and A_2 are situated at more than two steps from reward.

tion of these two categories is made in the tables.

Discussion

The results of the experiment just described permit us to conclude that there exists a phenomenon of spread in *recalling* success, comparable to the 'Thorndike effect.'

These two spread phenomena seem to be similar in so far as the punished *response* (Thorndike), as well as the *memory* of punishment, are both influenced by the proximity of a successful case. *The S reveals a tendency to conduct himself towards these failures as if they actually had been successes: repeating* them in a higher percentage of cases (Thorndike) and *recalling* them more frequently as having been rewarded.

On the other hand, there is a radical difference between the two phenomena, as regards the possibility of their theoretical explanation. The 'spread of effect' might be conceived eventually—as is done by Thorndike—as a *strengthening of connections* under the influence of a 'diffusion' (or a scattering) of the *confirmatory reaction* caused by the reward. However, if the reward strengthens *the neighboring S–R connections*, one would expect the connections between the response given and the result obtained to be strengthened too (S–R-punishment); i.e., the memory of the result obtained should be more true in the neighborhood of a rewarded response.

But what happens here is just the opposite. The memory of the punishment obtained is systematically falsified by the proximity of reward. Therefore, the phenomenon of *spread in recalling failure and success* requires, in any case, a different explanation from the one proposed by Thorndike for the *spread of effect*.

To throw further light on this question, the first point to be examined is whether the proximity of reward is effective during the first or rather during the second presentation of the series. Since the spread in recalling cannot be a matter of strengthening of connections, it may be expected, indeed, that the proximity of reward *during the learning experiment* (i.e., the first presentation of the list) is not essential.

With this end in view, a new series of experiments was set up.

Experiment II

The point in the method used in Experiment II is the shifting of the relative order of the punished connections during the second presentation of the series of words. The items immediately preceding or following reward in the first presentation of the series change places with the most removed cases during the experiment in recalling failure and success (second presentation).

The material is similar to that used in Exp. I; but the list contains 50 stimulus words instead of 40, and each word is followed by four nonsense words instead of five. The distribution of punishment and reward is fixed by a prearranged pattern, as indicated in Table I (Exp. II).

The procedure for the first part of the experiment is the same as in Exp. I. After the first presentation of the words, the *E* changes the instructions saying: "Listen to me now; in the meanwhile you must not look at your list." Saying this, he takes the list of words the *S* has in front of him, and lays it beside himself, behind a little screen placed on the table. Now, just as in Exp. I, he invites the *S* to repeat for each stimulus word read again, the *same* translation word he gave during the first presentation of the series, and to recall if his response was then said to be right or wrong.

After these instructions, the *E* gives back to the *S* his list saying: "Here is your list again; you may start now." Actually, the list given to the *S* is not the same. On this new list the relative order of certain items (i.e., stimulus words accompanied by the corresponding four nonsense words) is changed in the following way:

1. Punished cases situated between two rewarded items ('between two' cases) exchange places with items six or five steps removed from reward.

2. Items one step after or before reward exchange places with 'three or four step' cases.

3. The intermediate cases, namely the punished items two steps removed from reward, keep the same position in both orders. This is true of the rewarded cases, also; and the first two cases of the series always keep the same position.

It was observed that the *S*s did not become aware of these changes in the relative order of

TABLE IV

Wrong Responses Recalled as Rewarded After Reversal of Their Relative Position Towards Reward

The steps indicate the position of the wrong responses during the *second* presentation of the series.

Groups	*N*	Wrong responses changing places			For comparison:	
		1 step betw./aft./bef. (*N* = 2055) %	3 steps and more (*N* = 2055) %	Diff.	Right Responses %	Wrong Responses 2 steps %
W_1	30	24.9	21.8	+ 3.1	33.9	20.0
W_2	32	26.0	21.2	+ 4.8	34.1	25.8
W_3	24	27.2	15.0	+12.2	35.6	19.3
W_4	31	27.7	22.6	+ 5.1	31.1	27.8
W_5	20	33.7	27.7	+ 6.0	32.1	25.6
Total	137	27.5 SD = 1.1% SD (diff.) = 1.5%	21.5 SD = 1.0% (CR = 4)	+ 6.0		

certain items during the second presentation of the series.

This experiment was carried out with 137 high school boys (14 to 18 years old), divided into five groups (W_1 to W_5).

Results

The total results of Experiment II show that the proximity of reward is effective during the recalling trial of the experiment, and not during the first presentation (learning trial).

Comparing the recalls of the two classes of shifted items, i.e., the items closest to reward during the *first* presentation, with the items closest to reward in the *second* trial, we obtain the percentages shown in Table IV. It is stated there that the items in closest proximity to reward *during the second presentation* of the series, are recalled as rewarded in 27.5 percent of the cases. The responses which were farthest removed during this second presentation, but nearest during the first trial, obtain only 21.5 percent. This difference in percentages is similar to that obtained in Exp. I, and is statistically significant.

Therefore we conclude that the *Ss* reveal a slight tendency to consider more frequently as rewarded those items which, *during the recalling trial*, occur in the neighborhood of a previously rewarded response.

Spread in Recalling Failure—Experiment III

Still clearer insight into the mechanism of spread in recalling the results of serial activities is needed, before any theoretical hypothesis about it can be suggested. Therefore, a new type of experiment was devised, to examine the phenomenon in more simplified conditions. All learning task or memorizing effort will be avoided in this experiment, and only eight items will be presented to the *Ss*. The relative frequency of reward and punishment will be reversed: seven of the eight responses will be rewarded, one is to be punished. (Compare **14** and **20**.)

Procedure

The *S* is presented with eight cards (8 × 7 in.) in succession. Each card contains two thickly printed capital letters separated by a hyphen (T-J; Z-F; etc.). The *S* is asked to estimate by one glance at the card the relative size of the black surfaces of the letters on each card. He must say: "T is *larger* than J" or "T is *smaller* than J," beginning always with the first of the two letters (here: T). The *S* is told that this experiment is devised to examine the influence of shape in estimating surfaces.

A prearranged pattern for calling responses 'right' or 'wrong' was followed; all responses, except the estimate of the sixth card in the series, were called 'right'; the response on the sixth card was punished.

Sixty-four boys (12 to 16 years) took part in the experiment. They were divided into eight groups comprising eight subjects each. The relative order of presentation of the cards was changed for each group, so that another pair of letters was the punished item (the sixth card) in each of the eight groups. This was done to avoid any systematic influence of the nature of the letters on the recalls.

After he had completed his estimate in regard to all the cards, new instructions were given to the *S*. He was asked now to recall for each card, whether his estimate about the letters had been 'right' or 'wrong.' The cards were presented again in the same order as the first time. (The *S* however, is told nothing about this order.)

Results

The frequency with which the *right* responses are recalled as *wrong*, at the different steps of proximity to the punished item, is given in Table V. The total results of the eight groups, as indicated at the bottom of the table, distinctly reveal the influence of the proximity of the punished response upon the recalls. Summarizing, we obtain the following percentages:

one-step cases: 14.9 percent are recalled as wrong
two-step cases: 4.7 percent
3 steps and more: 1.6 percent.

In this experiment, the one-step cases *preceding* the punished item are favored with regard to the one-step cases *following* punishment (18.8 against 10.9).

The difference between the one-step cases and those on three steps and more is statistically significant.

After completing this experiment, some introspective data were obtained from the *Ss*. The *E* invited them to describe the general course of the experiment, and to tell him how many times their estimate concerning the size of the letters has been correct or wrong.

The following description is given, in general, by the *Ss*: "*In the beginning*, I made *several* correct estimates; but *near the end* (others say: *in the middle*) of the series, I made *one* failure." Seven *Ss* had the impression they made *two* failures, and one says he probably made *three* incorrect estimates.

Discussion

On the basis of the objective results and the introspective data obtained in Exp. III, we are invited to assume that the *Ss* 'keep in mind' a general impression of the whole series of succeeding rewards and punishments received. As a matter of fact, this impression may square only very approximately with the real configuration of what happened, especially when the series is more complicated than the simplified one we presented in Exp. III.

On the other hand, this assumption is corroborated by the following facts. The memory of the reward or punishment obtained in each individual case is very weak. As a matter of fact, the percentages of erroneous recalls of reward are very high for the different groups (cf. Tables II, III and IV). Nevertheless, the average percentage of items recalled as rewarded fits very well with the percentage of items really rewarded. For instance, the *Ss* of Groups A_1 (Exp. I) and those of Group W_1 (Exp. II) had respectively 35 and 24 percent of their responses

TABLE V

RESPONSES RECALLED AS PUNISHED AT THE DIFFERENT DISTANCES FROM THE PUNISHED ITEM AS STATED

(Total *N* of responses per group, for each step = 8.)

Groups	Right Responses Preceding the Wrong Response by					The Wrong Response	Right Resp. Following the Wrong Resp. by	
	5 steps (1st card)	4 steps (2d card)	3 steps (3d card)	2 steps (4th card)	1 step (5th card)	(6th card)	1 step (7th card)	2 steps (8th card)
F_1	0	1	0	0	1	4	0	0
F_2	0	1	0	1	2	2	2	0
F_3	0	0	1	0	4	4	1	1
F_4	0	0	0	0	2	5	1	0
F_5	0	0	0	0	0	6	3	0
F_6	0	0	0	1	2	7	0	0
F_7	0	0	0	1	0	7	0	1
F_8	0	0	0	0	1	5	0	1
Total	0	2	1	3	12	40	7	3
Percent	0.0	3.1	1.6	4.7	18.8	62.5	10.9	4.7

rewarded. In their recalls we find respectively an average of 40 and 25 percent of items recalled as rewarded (i.e., in each case a percentage very close to the objective percentages of 35 and 24), *although more than half these recalls are incorrect.*

Comparing the percentages of Groups A_1 and A_2 with those of the other groups of Exp. I, as indicated in Tables II and III, the same conclusion may be reached. The numbers of responses really rewarded amount here respectively to 35 and 20 percent; the percentages of responses recalled as rewarded approximate respectively 40 and 25. The same may be said about the only punished case in Exp. III. Therefore, we conclude that *the proportion of recalls is found to follow very closely the real proportion of actual failures and successes obtained in the course of the series, in spite of the fact that exact remembering is very weak.* This points to the fact that a *general impression* with regard to the results obtained is kept in mind and influences the recalls.

These facts confirm the hypothesis the present writer (10) proposed in a former paper on spread, on the basis of the results obtained with the first groups of Exp. I and Exp. II (Groups A_1 and W_1). This hypothesis may be completed now and described as follows.

Hypothesis: (a) Spread in Recalling

A very vague imprint or mark of the *general configuration of the series* seems to be kept in mind by a large number of *S*s. The factor responsible for establishing this configuration in the monotonous succession of items is the *contrasting result* obtained by a few responses. Experimental evidence is available at the moment, showing that a constrasting *result* (i.e., a failure in the middle of a series of successes or *vice versa*) has the same organizing effect as the contrasting *item* in a sequence of uniform material in the experiments of Köhler (4) and von Restorff (11). Even when no effort to recall is made during the first presentation of the series (Exp. III), the *S*s tell us they have from time to time the impression, during the recalling experiment, that about 'here' or about 'now' a success (or a failure, when the contrasting result is a failure) was obtained.

These facts find a confirmation in the experiments on anticipatory and so-called 'perseverative' responses, showing that a distinctive response in a serial learning task is often found to be erroneously localized *in the neighborhood of his right place* (5, 6, 13).

Therefore, it is assumed that the pattern of sub-groups, spontaneously formed by the contrasting results of some responses, may be a directive factor when, afterward, the *S* is trying to recall or to pick out the rewarded cases. In that way, the *S* will manifest a tendency to recall responses as rewarded after certain intervals (i.e., a certain number of intermediate items) corresponding approximately to the sub-groups of the series.

The factor described, however, is not the only one to influence the recalls of failure and success in serial learning. It was found in other experiments, not to be reported here (9), that the *degree of recognition* of an item is an important factor too. Many items presented during the first trial are not fairly recognized by the *S* during the second presentation of the series (recalling experiment). The *better recognized* situations, however, are more frequently and erroneously

recalled as having been successfully answered. Badly recognized situations, on the contrary, give a higher percentage of erroneous failure recalls.

It may be that situations immediately preceding or following a contrasting result (i.e., the first and the last item of a sub-group) are favored with respect to this point. In that case, the 'recognition factor' will interfere with the former in producing the spread of failure recalls and the two factors will cooperate with each other in producing erroneous recalls of success.

Hypothesis: (b) Spread of Effect

Discussing the possibility of explaining learning by the *informative* influence of reward, Thorndike said (**16**, pp. 29–30):

> The imformative influence of a reward would not be expected to spread or scatter to strengthen neighboring wrong connections. . . . A secondary image or idea of 'Right' derived from the informative influence of a reward would have not only to get loose from the 'word-number-Right' sequence to which it was attached, but also to knock out the third member of some 'word-number-Wrong' sequence and take its place.

From the description of our experimental results on spread in recalling success, it appears that the "informative influence of reward" in what is called the *spread of effect* may eventually be considered in a less mechanical way no third member of a sequence is to be knocked out, as Thorndike thought; as a matter of fact, the eventually 'informative' recalls of success are found to spread or scatter around the really rewarded response. On the other hand, it was found that in serial learning *reward as such* does not strengthen even the rewarded connections (**8**, **18**). In that way, reward cannot be expected to *spread* or *scatter* any strengthening effect.

Therefore, it is suggested that there may be some relation between the slightly higher percentage of *repetitions of wrong responses* in the proximity of reward (Thorndike-effect) and the slightly higher percentage of *recalls or reward among these punished cases.*

As a matter of fact, it was ascertained in Exp. I and Exp. II that the Ss were *not able to reproduce* more correctly the punished responses in the neighborhood of reward than the more removed ones. During the second presentation of the series of items (recalling experiment) an average of 50 percent correct reproductions was obtained for the wrong responses two steps and more removed, as well as for the 'one step' cases.

On the basis of this fact, we conclude that the S–R connections are not really *strengthened* or better *learned* by the very fact of proximity of reward; they only are more frequently *repeated* in the course of a *learning experiment.* The *motivation* of that higher percentage of repetitions may precisely be found in the slight tendency to recall these responses more frequently as rewarded.

This influence of recalls on the repetition of responses must not necessarily be conceived as an 'informative influence,' as if a deliberate choice was to be made between several responses available, on the basis of the *knowledge* that one of them has been right. We agree with Thorndike that this is not likely to happen in these experiments. But *a less explicit 'valence' may be attached to responses in the proximity of reward, by reason of the vague impression of success in that region.*[2]

[2] It is important to notice that with regard to many items the Ss recall the response they gave during the first presentation of the series, without remembering if that response was said to be 'right' or 'wrong.' (See Dand, A. Reward and

Moreover, it is not claimed that the new factor of spread in recalling success is to be considered as 'the' explanation of the Thorndike-effect. In the opinion of the present writer, different factors may be effective in producing this phenomenon. The phenomenon of spread in recalling success is not found to exist in the experimental results of each of our *Ss*; in others the guessing-sequence hypothesis of Jenkins and Sheffield (3) or the recognition factor (9) may be rather effective. They may account also for the experimental results obtained by Muenzinger and Dove (7). Several factors may eventually interfere or interact in the same *Ss*. Most of these factors are nothing but slight subjective tendencies, which may depend upon variable attitudes in the *Ss*; they all may cooperate in certain ways in producing the slight and irregular 'effect,' that does not seem to have the importance which it was expected to present for a theory of learning. This decrease in the theoretical importance of the Thorndike-effect is the most striking conclusion to be drawn from the recent experiments in that field.

Conclusions

1. Experimental evidence is given to show that a phenomenon of spread, analogous to the 'Thorndike-effect,' exists among the *recalls* of failure and success.

2. The spread of recalls, however, cannot be conceived as a strengthening of connections. On the contrary, it implies a substitution of the previously connected 'S–R-punishment' sequence, by the falsified sequence 'S–R-reward.'

3. The proximity of reward is effective (in producing the spread of recalls) during the second trial (i.e., during the recalling experiment), and not during the first presentation of the series (learning experiment).

4. It is suggested by the experimental results, as well as by some introspective data, that the *Ss* keep in mind a vague impression of the general configuration of the series. The monotonous series of punished responses is divided up in sub-groups by the *contrasting rewards*. This vague imprint may be a directive factor in recalling failure and success; it seems to be responsible for the fact that a slightly higher percentage of responses is recalled as rewarded in the neighborhood of a previously rewarded case.

5. It is suggested that the spread in recalling success may have some influence on the fact that wrong responses are more frequently *repeated* in the proximity of a response previously rewarded (spread of effect).

Summary

Experiments were set up to examine whether punished responses in the proximity of a successful one are more frequently *recalled* as 'rewarded' than are more remote items (*spread in recalling success*).

In a first series of experiments, with 12 groups of *Ss*, it is found that a phenomenon of 'spread in recalling success' really exists.

A second series of experiments shows that the factor responsible for the spread of recalls is the proximity of reward *during the second presentation of the series*, i.e., during the recalling experiment, and not during the first learning trial.

---

punishment in learning. *Brit. J. Psychol.*, 1946, 36, 83–87.) In that way, there may be a tendency to repeat these responses in the neighborhood of success and to avoid them in other zones of the series.

In a third experiment, the mechanism of spread is examined in simplified conditions. It is shown that recalls of *failure* spread as well as recalls of success, when punishment is the *contrasting* result obtained by the *Ss*.

It is suggested by the experimental results and by some introspective data that the monotonous series of punished responses is divided up in sub-groups by the contrasting rewards, and that a vague imprint of this succession of sub-groups is kept in mind.

An hypothesis is built up, on the basis of these facts, to explain the new phenomenon of spread in recalling failure and success. The relation between this phenomenon and the Thorndike-effect is shown.

(Manuscript received
December 13, 1948)

References

1. Donceel, J. *Le rôle du souvenir dans l'interprétation de la loi de l'effet.* Unpublished thesis under the direction of Prof. A. Michotte. Univ. Louvain, 1934.
2. Farber, F. Spread of effect of reward and punishment in a multiple choice situation. *Proc. Iowa Acad. Sci.*, 1941, **48**, 313–317.
3. Jenkins, W. O., & Sheffield, F. D. Rehearsal and guessing habits as sources of the 'spread of effect.' *J. exp. Psychol.*, 1946, **36**, 316–330.
4. Köhler, W., & von Restorff, Hedwig. Analyse von Vorgängen im Spurenfeld. I. Über die Wirkung von Bereichsbildung im Spurenfeld. *Psychol. Forsch.*, 1933, **18**, 299–342.
5. Lumley, F. H. Anticipation of correct responses as a source of error in the learning of serial responses. *J. exp. Psychol.*, 1932, **15**, 195–205.
6. Lumley, F. H. Anticipation as a factor in serial and maze learning. *J. exp. Psychol.*, 1932, **15**, 331–342.
7. Muenzinger, K. F., & Dove, C. C. Serial learning. I. Gradients of uniformity and variability produced by success and failure of single responses. *J. gen. Psychol.*, 1937, **16**, 403–413.
8. Nuttin, J. Respective effectiveness of success and task-tension in learning. *Brit. J. Psychol.*, 1947, **38**, 49–55.
9. Nuttin, J. *The influence of the recognition of a situation on the impression of success.* (Unpublished)
10. Nuttin, J. De finaliteit in het menschelijk handelen en het connectionnisme. Een studie nopens de wet van het effect. *Tijdschrift voor Philosophie*, Louvain, 1942, **4**, 235–268.
11. Nuttin, J., & Degadt, O. *Strengthening of connections by a contrasting result.* (Unpublished)
12. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
13. Spragg, S. D. S. Anticipatory responses in serial learning by chimpanzee. *Comp. Psychol. Monogr.*, 1936, **13**, 1–72.
14. Stone, G. R. The effect of negative incentive on serial learning. I. The spread of variability under electric shock. *J. exp. Psychol.*, 1946, **36**, 137–142.
15. Thorndike, E. L. An experimental study of rewards. *Teach. Coll., Columbia Univ., Contr. Educ.*, No. 580. New York: Teachers College, Columbia Univ., 1933.
16. Thorndike, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton-Century, 1935.
17. Tilton, J. W. Gradients of effect. *J. genet. Psychol.*, 1945, **66**, 3–19.
18. Wallach, H., & Henle, Mary. An experimental analysis of the law of effect. *J. exp. Psychol.*, 1941, **28**, 340–439.
19. Wallach, H., & Henle, Mary. A further study of the function of reward. *J. exp. Psychol.*, 1942, **30**, 147–160.
20. Zirkle, G. A. Success and failure in serial learning. II. Isolation and the Thorndike effect. *J. exp. Psychol.*, 1946, **36**, 302–315.

THE PERCEPTION OF THE VERTICAL: II. ADAPTATION EFFECTS IN FOUR PLANES[1]

BY GEORGE E. PASSEY AND FREDERICK E. GUEDRY, JR.

Department of Psychology, Tulane University of Louisiana and School of Aviation Medicine and Research, Naval Air Station, Pensacola, Florida

I. Introduction

The airman's estimation of straight and level flight is often affected severely by sensory illusions, and false judgments of spatial position are common occurrences in flight. That man's senses may give rise to perceptions of spatial orientation which are at variance with actual spatial position has been noted by many investigators. It is possible that sensory illusions may affect the pilot's alignment of the axes of an aircraft with the axes of space with regard to both magnitude and direction of error. For example, he may correctly detect the amount of inclination in pitch or roll while erroneously detecting the direction of inclination, or the converse might be true.

A deviation in alignment of the vertical axis of the aircraft with the gravitational vertical is a function of (a) the pilot's estimate of 'straight and level,' and (b) the pilot's ability to maintain what he considers 'straight and level' flight. If the pilot's estimate of 'straight and level' is incorrect, these two axes can coincide only when the pilot is unable to maintain what he considers a 'straight and level' attitude or when the aircraft is exactly inverted. Some of the commoner illusions have recently been reviewed by Vinacke (12).

The main determinants of spatial orientation are agreed by most investigators to be of visual and postural origin. Gibson and Mowrer (5) defend the notion of the supremacy of postural factors. However, Asch and Witkin (2) indicate that visual cues are the primary means of spatial orientation. Where visual indicator cues are present, they may be readily substituted for the normal visual field; on the other hand, it is hardly possible to present postural indicator cues which could be substituted for the normal complex of postural cues. Use of both postural and visual cues is of vital importance to the aviator. Investigation of postural cues alone followed by the addition of visual cues to note any improvement in performance should be of help in clarifying this issue. The present report will deal, however, with the use of postural cues alone.

Aviators recognize that when the normal visual framework or its representative on the instrument panel is absent it is possible to assume many flight positions without perceiving the change in attitude of the aircraft. With postural cues operating alone, deviation from 'straight and level' flight often is not detected as a result of adaptation or inadequate stimulation. Burtt (3) has reported that rapid inclination from the gravitational vertical in the absence of a visual field gives rise to detection of tilt for smaller magnitudes of tilt than when a less rapid movement is used. This suggests that inadequate stimuli rather than an adaptation effect produce a certain amount of disorientation. Jones,

[1] This report is part of a series of investigations conducted jointly with the School of Aviation Medicine and Research under Contract N7onr-434, Task Order I, with the Office of Naval Research in cooperation with the Bureau of Medicine and Surgery and the Tulane University of Louisiana. Project designation number NR140-455 of the Medical Sciences Division, Office of Naval Research. Project NM-001-037 of the Research Division, Bureau of Medicine and Surgery. Task Order Director: Cecil W. Mann.

Milton, and Fitts (7) found that their pilots could not detect inclination from the gravitational vertical adequately enough for pilot safety. MacCorquodale (9) and MacCorquodale, Graybiel, and Clark (10) also noted this inability when they increased head to seat force. Increase in this force without change in attitude of the aircraft was interpreted as a climb. Backward tilt was reported in turns in which the angle of attack remained constant while power setting was increased.

Gibson and Radner (6) and Gibson (4) note adaptation in the adjustment of a luminous line to the vertical when this line is not anchored to a visual frame of reference. They point out that within the limit of the times investigated, zero to 120 sec., exposure to the tilted lines produced adaptation effects as evidenced by progressively greater errors in readjusting the line to the gravitational vertical as the exposure time increased.

Most of the investigators reporting in this field have found evidence which seems to indicate adaptation. Aircraft pilots report that a gradual departure from 'straight and level' flight with maintenance of inclination for some period of time leads to a feeling of 'straight and level' in the tilted position. When a subsequent return to 'straight and level' is made, bank and turn in the opposite direction is often reported. Whether or not this apparent adaptation will be reported by subjects placed in a device which is free to tilt in any direction but which does not reproduce any of the other motions of an aircraft is a question worthy of investigation. The primary purpose of this study is to determine whether or not this apparent adaptation effect is indicated by subjects placed in such a device. Comparison of results obtained when adjustment to the vertical is attempted immediately after inclination with those obtained when the individual is exposed to tilt for a period of time prior to adjustment to vertical may indicate adaptation, if any be present. The following hypotheses will be investigated in this study:

1. In the absence of a visual frame of reference, subjects maintained in tilted positions prior to adjusting to vertical will adjust as accurately as when they make their adjustments immediately after being tilted.

2. In any plane investigated in this study, the two possible directions of tilt (e.g., fore tilt and aft tilt in the medial plane) will not have a differential effect upon the subject's adjustment to the vertical.

3. Adjustments to the gravitational vertical from a titlted position are made equally accurately in all planes which pass through the gravitational vertical.

II. Apparatus and Procedure

A C3 Link Trainer was the principal item of equipment used in this experiment. The trainer was capable of inclination to a maximum of approximately 18° in any plane passing longitudinally through the gravitational vertical. Removal of the turning bellows and a number of the automatic features allowed pitching and rolling of the trainer without change in heading. The trainer was mounted on a universal joint and supported by four bellows which could be used to maintain or change the attitude of the trainer. In 'straight and level' flight position, the gravitational vertical and the subject's body axis were in alignment, so that the imaginary axes both passed through the universal joint. The control stick was the only control left for use by the *S*. Movement of the stick in any direction moved the trainer in that direction. The mechanism was so arranged that one of a pair of bellows received atmospheric pressure while its companion was evacuated causing the trainer to move in the direction of the evacuated bellows. *E* was provided with two manually operated valves, which by-passed the subject-operated valves and allowed the *E* to move the trainer independently.

The inclination of the trainer was recorded simultaneously in the medial plane and in the lateral plane by two independent selsyn receivers. Transmitters were mounted with their

shafts parallel to the longitudinal and transverse axes of the trainer. The fields of the transmitters were firmly attached to and rotated with the trainer. The armatures of the transmitters retained a constant position by virtue of a pendulum attached to each armature shaft. All of the inclinations were determined by resolving the two inclinations.

Sixteen *Ss* were used. Experimental group I consisted of eight *Ss* with previous experience in psychological experiments; their familiarity with this and like problems was recognized. Experimental group II consisted of eight volunteers from among the flight student group at the Naval Air Station, Pensacola, Florida. Members of this group had not had previous experience in link trainers, aircraft, or in psychological experiments.

The *Ss* were successively tilted in four planes. These planes passed through the vertical axis of the trainer and the gravitational vertical and were designated with respect to the head of the trainer as (1) medial (0–180 degrees); (2) left oblique (135–315 degrees); (3) lateral (90–270 degrees), and (4) right oblique (45–225 degrees). Two tilt positions were used in each plane, each in opposite directions from the other. In the case of experimental group I, the inclinations in the medial and lateral planes were of approximately 10 degrees, and those of the oblique planes were approximately 14 degrees. All of the inclinations of experimental group II were of approximately 10 degrees magnitude.

Of the four planes described above, two were investigated on one day and the other two the following day, so that each *S* was tilted in all four planes. One session occurred each morning and one each afternoon. Two different planar orders were used. Four *Ss* in each experimental group received the order 1–2–3–4, and constituted the first sub-group, while the second four received the order 4–3–2–1, and constituted the second sub-group. Two *Ss* of each sub-group received the inclinations in R-L-L-R-L-R-R-L-R-L-L-R-L-R-R-L-R-L-L-R order (R = right and L = left), while the second two *Ss* of each sub-group received the order L-R-R-L-R-L-L-R-L-R-R-L-R-L-L-R-L-R-R-L. In the medial plane these orders became F-A-A-F-A-F-F-A-F-A-A-F-A-F-F-A-F-A-A-F and A-F-F-A F-A-A-F-A-F-F-A-F-A-A-F-A-F-F-A (F = forward, and A = aft). These orders were associated in turn with (1) an immediate adjustment to vertical following inclination (condition of no delay), and (2) adjustment to vertical after being in a tilted position for 60 sec. (condition of delay).

Twenty determinations were made in each plane. The determinations were made in units of 10 separated by approximately one-half hour rest periods. Ten determinations were made from inclinations in each direction. Five inclinations in each direction were made with delay and five with no delay.

Ss were brought into the experimental room, placed in the trainer, and given instructions as to operation of the trainer and the experimental procedure.

Following three practice trials with reference to visual objects, goggles were lowered, which eliminated the visual field, and the experimental series was given. At the start of each subsequent series of determinations the trainer was unlocked and *S* was allowed to bring the trainer to 'straight and level' flight using visual cues. Following this, *S* lowered his goggles for administration of the experimental series.

III. Results [2]

Our first hypothesis stated that adjustments to 'straight and level' following exposure to inclination for a period of approximately 60 sec. would be made as accurately as when *S* was allowed to readjust the trainer immediately following his inclination. Mean error in degrees was computed for non-delay readjustment (ND) and delayed readjustment (D). The means represented the mean of 10 adjustments to the vertical and were computed for each subject in each plane. The means were compared for each experimental group. The mean error for experimental group I under a condition of non-delay was 2.7 degrees and that for delay was 4.4 degrees. The t computed for the difference between these means was 10.62. The mean error for experimental group II under a condition of non-delay was 2.8 degrees, and that for delay was 4.4 degrees. The

[2] Tables containing the basic data from which the comparisons were made are not presented herewith because of space limitations. These tables can be obtained by ordering Document 2669 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $0.50 for microfilm (images one in. high on standard 35 mm. motion picture film) or $0.50 for photocopies (6 × 8 in.) readable without optical aid.

t computed for the difference between these means was 7.85. The degrees of freedom for both groups were 31 and the $P < .001$.

The basic data were examined and another approach to check the first hypothesis was made by tabulation of the number of errors in the direction of initial inclination for each plane. The data were again grouped for planes and conditions and the mean number of errors in the direction of initial inclination were compared by means of the t statistic for each experimental group. For experimental group I the mean number of errors in the direction of initial inclination under a condition of non-delay was 4.7 and that for delay was 6.9. The t computed for the difference between these means was 5.76. For experimental group II the mean for non-delay was 5.3 and the mean for delay was 7.3. The t computed for the difference between these means was 4.55. The degrees of freedom were 31 and the $P < .001$.

Use of these two indices allows rejection of the first hypothesis. Following exposure to tilt for a period of 60 sec., readjustments to the gravitational vertical are significantly less accurate than under immediate readjustment, and the number of errors in the direction of initial inclination is significantly greater under a condition of exposure to inclination giving evidence of adaptation.

Comparison of the two experimental groups with respect to the absolute divergence of their settings from the gravitational vertical yielded a t of .21 ($df = 62$). There is a probability between .9 and .8 that such a difference occurred by chance and hence there is no reason to suppose that the two groups differed with respect to performance of this task.

When considering variability of the data under immediate and delayed resetting, it was found that there was significantly more variability for delayed than for immediate resetting. An F of 2.34 was obtained (df for each variance = 63, $P < .01$).

The second hypothesis stated that adjustments would be made equally accurately in approaching the vertical from either direction of inclination in a single plane. Data for immediate and delayed readjustments were combined, and the analysis was made with respect to the direction of inclination in the plane involved. The t's which allow comparison are presented in Table I. None of the differences was significant. On the basis of our findings, we accept the hypothesis as stated.

Our third hypothesis that setting would be made equally accurately for any direction of inclination may be

TABLE I

THE t-VALUES FOR AVERAGE ERROR IN TWO DIRECTIONS OF INITIAL INCLINATION IN EACH PLANE

Subjects	Planes							
	Medial		Right Oblique		Lateral		Left Oblique	
	t	P	t	P	t	P	t	P
Group I	.81	>.4	1.23	>.2	.67	>.5	.56	>.5
Group II	.97	>.3	.88	>.4	1.18	>.2	1.76	>.1

examined by computing the mean error in degrees for the 20 adjustments made by each subject for each plane. Although the initial inclinations for the two experimental groups differed in the oblique planes by four degrees, the difference in their performance is not significant, and, for the purpose of this analysis, the two groups are considered together. A comparison was made, and the t's obtained are presented in Table II.

TABLE II

INTERPLANAR t-VALUES ($df = 15$)

Planes	t	P
Lateral vs. R. Oblique	4.32	<.001
Lateral vs. L. Oblique	4.09	<.001
Lateral vs. Medial	3.05	<.01
Medial vs. R. Oblique	1.74	>.10
Medial vs. L. Oblique	1.56	>.10
L. Oblique vs. R. Oblique	1.03	>.30

Performance in the lateral plane differs from performance in all other planes investigated, it being the most accurate plane. When lateral is compared with either oblique plane, the $P < .001$, and when compared with the medial plane, the $P < .01$. The medial planes give some evidence of being superior to the oblique planes, the $P < .20$, while the oblique planes show little probability of superiority of one over the other, $P < .40$. On the basis of these results, we may reject the hypothesis as stated.

IV. Discussion

The data obtained allow rejection of the first stated hypothesis at the .001 level of confidence. These data confirm the work of Jones, Milton, and Fitts (7), MacCorquodale (9), and MacCorquodale, Graybiel, and Clark (10) on adaptation during the prolonged turning of aircraft. It must be remembered, however, that our findings show adaptation when the individuals are exposed to a force not in alignment with the body axis. It is probable that after the accelera-tory stimuli involved in movement have ceased to operate and the force is aligned with the body axis with no visual framework present, the individual would have much less to go on in the making of a judgment than in the present experiment. His adjustment in the earlier experiments would have to be made in great measure upon the increased force along the body axis while in our experiment divergence of the two axes is a cue. As the individual returns to verticality, in our experiment, the force on the body axis increases and judgment is probably made in part on this basis.

Comments by *Ss* in this experiment indicated that judgments following 60 sec. of exposure to tilt were more difficult to make than those involving immediate resetting. On occasional trials, some *Ss* from the fiight student group stated that during the delay they felt they were gradually being returned to 'straight and level.' Following delay and upon being instructed to return to 'straight and level,' some failed to move the trainer at all and in a few instances increased the degree of tilt in the same plane and direction. Pilots in other studies following a turn involving some delay reported that they were turning in the opposite direction.

There is greater variability shown for the data involving delayed readjustments. The F computed is significant at the one percent level of confidence. Mann, Berthelot-Berry, and Dauterive (11) have proposed the concept of the 'arc of uncertainty,' $U°$. They noted a significant increase in the size of this arc following a reduction in proprioceptive cues, achieved by use of a softly padded

seat in one instance in contrast to use of a seat without padding in another. The data of the present experiment show a significant increase in the arc of uncertainty when a delay is introduced. Since no delay was introduced in their experiment, we have here another basis for increased variability.

Mann, Berthelot-Berry, and Dauterive also found a consistent tendency toward underestimation in returning to vertical. When the trials involving immediate resetting are considered it is noted that of the 640 determinations made, 319 were in the direction of initial tilt, and 321 were in the direction of overestimation. The difference is not significant. When, however, the 640 determinations made under delay are examined, it is noted that 463 of these settings were made in the direction of initial inclination or were underestimations, while only 177 were overestimations of the vertical. A t computed here indicates a $P < .01$. Underestimations of the amount of tilt are characteristic when *Ss* are returning to vertical following exposure to tilt for 60 sec.

The third hypothesis under investigation proposed that judgments would be equally accurate in all of the planes passing through the gravitational vertical which were investigated in this study. Burtt (3) found in an early study that in the absence of a visual field, individuals are better able to recognize the gravitational vertical when they are tilted in the lateral plane than when tilted in the medial plane. Later, Jones Milton, and Fitts (7) found similar results in aircraft. They determined that aviators made more errors in judgments of pitch than they did in judgments of roll. Their subjects were limited in the kinds of errors in that they made only verbal reports. It must be remembered, however, that forces in the latter study acted from head to seat, and those in Burtt's study had the principal force removed from head to seat. Our data lend support to their findings and in addition allow comparisons in other planes.

The factors operating in this result are not clear at this time. Increase in variability found by Mann, Berthelot-Berry, and Dauterive following reduction of the proprioceptive cues gives evidence of the dependence upon bilateral stimulation in lateral tilt. If accuracy depends upon the availability of proprioceptive cues, then an individual tilted aft in the oblique planes should have additional stimulation from the back of the seat plus the bilateral stimulation of lateral tilt. The results we have do not bear this out.

In everyday situations, we are used to being tilted in the medial plane. When we sit in a comfortable chair, or when we bend over a desk, we displace our body axis from the gravitational vertical in the medial plane. We very rarely, however, displace our body axis from the gravitational vertical laterally. We seldom walk with a list. We may detect inclinations laterally more easily since we are unaccustomed to them.

While the emphasis in our discussion has stressed the proprioceptive cues, labyrinthine stimulation must not be ignored. A study by Lewis and Horn (8) indicated that deaf-mutes are unable to detect tilt and motion of aircraft. Arndts (1), using surface anesthesia in the absence of a visual field, emphasized the importance of cues of 'deep sensibility.'

The establishment of adaptation to inclination in this experiment and the casual references to this phenomenon by other investigators in flight situations indicates that it is an im-

portant phenomenon in pilot disorientation. This study shows no evidence of improvement from session to session. Pilots, however, do improve with experience with the detection of slips and skids of an aircraft. Since we find no improvement, it may be that better perception of skids and slips by the experienced pilot involves attention to these cues as the other complexities of flight require less of his time. The factor of rate and magnitude of displacement of the vector from the body axis may be highly important factors. Further investigations of the phenomenon of adaptation are indicated.

V. Summary

1. Sixteen *S*s were required to return a modified Link trainer to 'straight and level' flight following inclination.

2. Inclinations were made in four planes passing longitudinally through the gravitational vertical and were designated as medial, left oblique, lateral, and right oblique. Twenty adjustments were made in each plane during four experimental sessions in which each session was devoted to making determinations in one plane only. Ten adjustments were made from each direction within a plane and half of the trials in each direction were made under conditions of immediate return to the vertical after inclination while half of the trials entailed a 60-sec. delay before readjustment was allowed.

3. When either absolute divergence from the gravitational vertical in degrees or number of errors in direction of initial inclination is considered, there is a significant adaptation effect.

4. Delay also serves to produce significant difference in variability.

5. There is no significant difference in performance shown in comparison of adjustments from different directions of inclination in the same plane.

6. Adjustments in the lateral plane are made with greater accuracy than adjustments in either medial, left oblique, or right oblique. The differences are significant at the one percent level.

(Manuscript received October 29, 1948)

References

1. Arndts, F. Zur Frage nach den der Lagewahrnehming dienenden Sinnesfunctionen. *Pflüg. Arch. ges. Physiol.*, 1924, **204**, 539–540.
2. Asch, S. E., & Witkin, H. A. Studies in space orientation: I. Perception of the upright with displaced visual fields. *J. exp. Psychol.*, 1948, **38**, 325–337.
3. Burtt, H. E. The perception of slight changes of equilibrium with especial reference to problems of aviation. *J. appl. Psychol.*, 1918, **2**, 101–115.
4. Gibson, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast in the areal restriction of the after-effect. *J. exp. Psychol.*, 1937, **20**, 553–569.
5. Gibson, J. J., & Mowrer, O. H. Determinants of the perceived vertical and horizontal. *Psychol. Rev.*, 1938, **45**, 300–323.
6. Gibson, J. J., & Radner, M. Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies. *J. exp. Psychol.*, 1937, **20**, 454–467.
7. Jones, R. E., Milton, J. L., & Fitts, P. M. An investigation of errors made by pilots in judging the attitude of an aircraft without the aid of vision. Aero. Med. Lab., Eng. Div., Wright Field. TSEAA-694-13. May 2, 1947. Pp. 22.
8. Lewis, E. R., & Horn, H. Medical studies on the "feel of the airship." Deaf-mutes and normals. *Laryngoscope*, 1919, **29**, 63–81.
9. MacCorquodale, K. The effects of angular acceleration and centrifugal force on non-visual space orientation during flight. I. Methodology and preliminary results. Bu. Med. Research Proj. X-148 (Av-4-3),

Rept. 14. School of Aviat. Med. & Research, Pensacola, Fla. 13 July 1946. Pp. 12.

10. MacCorquodale, K., Graybiel, A., & Clark, B. The effects of angular acceleration and centrifugal force on non-visual space orientation during flight. II. Influence of habituation and technique of assuming the turn. Bu. Med. Research Proj. X-148 (Av-4-3), Rept. 15. School of Aviat. Med., Pensacola, Fla. 22 July 1946. Pp. 8, iv.

11. Mann, C. W., Berthelot-Berry, N. H., & Dauterive, H. J. The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.

12. Vinacke, W. E. Illusions experienced by aircraft pilots while flying. *J. Aviat. Med.*, 1947, **18**, 308–325.

PERMANENCE OF EXPERIMENTALLY INDUCED CHANGES IN THE ATTRACTIVENESS OF ACTIVITIES

BY MILDRED E. GEBHARD

University of Pennsylvania

In two recent studies (1, 2) expectations and experiences of success and failure, experimentally varied, were shown to be accompanied by changes in the attractiveness of the experimental task. The more recent of these studies was designed to yield supplementary information on the permanence of these experimentally induced changes in attractiveness; specifically, the permanence of such changes over the period of one day.

It can be hypothesized that the permanence of these observed changes in attractiveness is contingent upon the continuing effectiveness of the same factors which were originally associated with the changes. Previously reported analyses of the dynamics of the experimental situation and of the obtained data have both supported the hypothesis that *feelings of success and failure*[1] were predominantly responsible for the increases or decreases in the attractiveness of the performed activity. Hence, in a consideration of the permanence or lack of permanence of these observed changes in attractiveness, the permanence or lack of permanence of these feelings of success and failure is of primary importance.

A brief, schematic analysis of the previous experiment appears in Table I. The experimental variables were (a) pre-performance expectation of success or failure and (b) experience of success or failure, which, in factorial design, make a total of four experimental conditions (Columns 1 and 2). *Expectations* of success and failure were experimentally varied by varying the apparent difficulty of the experimental task. Half of the *S*s had reason to believe that chance had favored them with an unusually easy set of problems; the other half, that chance variation had resulted in their obtaining an unusually difficult set (Column 3). *Experience* was experimentally varied by reporting the performance to stand at either the 80th or 20th percentile of a comparable college population (Column 4). It was predicted that *S*s succeeding on a difficult task would experience greater success than *S*s succeeding on an easy task; similarly, that *S*s failing on a difficult task would experience less failure than *S*s failing on an easy task (Column 5). Changes in the attractiveness of the performed task were predicted (Column 6) and did follow (Column 7) in the rank order of the predicted intensities of feelings of success or failure.

It is hypothesized (1) that over the period of one day these *feelings of success* remain relatively unchanged but that *feelings of failure* are modified.[2] The concept of success, whether employed in a general or a specific sense, connotes satisfaction of the needs or desires of the organism. Feelings of success, concomitent with this satisfaction, demand no modifica-

[1] The term 'feeling' was employed to point up the distinction between subjectively experienced success and failure and operationally defined experience of success or failure. It is not employed here in the strict and limited sense of affective experience.

[2] It must be recalled that the experimental procedure was, by design, conducive to limited ego-involvement, i.e., little need to be successful or to avoid failure in the experimental situation. No generalization to strongly ego-involving situations is intended or attempted here.

tion. By contrast, failure, in both general and specific senses, connotes lack of satisfaction or frustration of the organism, and is followed almost inevitably by ideational or behavioral responses which tend toward satisfaction directly or indirectly. The adequacy of these responses can sometimes be determined through direct observation; on other occasions it is inferred, as by noting the extent to which the original feelings of failure appear to be moderated or dispelled.

It is further hypothesized (2) that these feelings of success remain relatively unchanged despite variation in the circumstances accompanying the success, but that feelings of failure are modified, more or less readily, as a function of the circumstances which accompany the failure. Examination of the experimental situation indicates that both success and failure are, under one condition, capable of interpretation as the result of personal adequacy or inadequacy, and, under another condition, capable of interpretation as the result of situational and uncontrollable factors. These circumstances have already been acknowledged in the postulation of varying degrees of experienced success and failure (Column 5), and are further hypothesized to relate to the modifiability or permanence of these feelings, as follows. (a) Feelings of success arising from a situation in which good performance is a function of the favorable operation of uncontrollable situational factors (e.g., good luck) do not need and probably will receive no modification. Witness the enduring elation engendered by success in games of chance. (b) Feelings of success arising from a situation where good performance is a function of personal ability do not need and probably will receive no modification. (c) Feelings of failure arising from a situation in which poor performance is a function of personal inadequacy need modification but are not subject to easy or early modification. This is especially true if no immediate direct or substitute satisfactions are available. (d) Feelings of failure arising from a situation in which poor performance is a function of the unfavor-

TABLE I

SCHEMATIC ANALYSIS OF PREDICTED PERMANENCE OF CHANGES IN THE ATTRACTIVENESS OF PERFORMED ACTIVITIES ACCOMPANYING EXPERIMENTAL VARIATION OF EXPECTATION AND EXPERIENCE

1	2	3	4	5	6	7	8	9
Experimental Condition		Nature of Task (Chance)	Level of Achievement	Predicted Actual Feelings	Predicted Changes in Attractiveness	Average Changes in Attractiveness*	Predicted Feelings One Day Later	Predicted Permanence of Changes in Attractiveness
Expectation	Experience							
Success	Success	Easy	80 %ile	Success	Rise	+1.4	Success	No change
Failure	Success	Difficult	80 %ile	*Success*	*Rise*	+1.7	*Success*	No change
Success	Failure	Easy	20 %ile	*Failure*	*Fall*	−1.4	*Failure?*	Rise
Failure	Failure	Difficult	20 %ile	Failure?	Fall?	+0.7	Failure??	*Rise*

* On a nine-point scale from a midpoint of five

able operation of uncontrollable situational factors (e.g., bad luck) will very likely be moderated or even completely dispelled as time provides further opportunity for fuller recognition of the origin of failure and of the adequate grounds for personal exoneration from responsibility for the failure (Column 8).

(3) It is hypothesized that the permanence or lack of permanence of these previously observed changes in attractiveness is a function of the permanence or lack of permanence of feelings of success or failure. This appears to be a legitimate extension of the original hypothesis that the changes in attractiveness of activities were a function of the feelings of success or failure, the rationale of which and evidence in support of which have been presented previously (2).

(4) Finally, it is hypothesized that with moderation of feelings of failure the attractiveness of the activities will rise. Support for this derives from the previous experiment in which increases or decreases in attractiveness were indicated to be a function of the intensity of feelings of success or failure.

From these considerations have been derived the following predictions: (1) That the rankings of attractiveness made under the condition of experienced success will not vary significantly over the period of a day; (2) That the rankings of attractiveness made under the condition of experienced failure will rise, but that the rise will be greater under the condition of expected failure, experienced failure than under the condition of expected success, experienced failure.

Experimental Design and Procedure

The experimental design and procedure of this study have been reported in detail elsewhere. In brief, 40 women college students ranked a series of nine tasks in order of attractiveness, and each performed the task to which she had given the median rank of five. Pre-performance expectations of success or failure and experiences of success or failure were experimentally varied in a factorial design. Consequent changes in attractiveness were determined by noting the position of the performed task in a second ranking of the tasks. At the conclusion of this procedure *E* stated that there was insuificient time for the performance of a second task, and asked each *S* if she would be willing to do another task at home in her leisure. *S* was asked to come back the following day to pick up the task which would be ready for her then.[3] The procedure of the second day has not been reported, and is indicated by the following instructions to the *S*s.

> Before I give you the task you are to take with you I want to ask several questions about the tasks we used yesterday. Sometimes when you have been away from tasks for a while you change your mind as to how much you like to do them. I want you to do now as you did before, list the tasks in the order in which you like to do them. I don't care whether or not your ratings are consistent with those you made yesterday. In fact, it will be best if you don't try to remember what you did yesterday but let your ratings indicate exactly how you feel about the tasks now.

The data studied in this paper consist of these third rankings of attractiveness.

Results and Discussion

Nine tasks were ranked in order of attractiveness three times by each *S*. The task given the rank of five in the first ranking (i.e., the critical task) was performed by the *S* and changes in its attractiveness were determined by observing its position in the two later rankings. The second ranking, made immediately after the performance of this task (i.e., the immediate ranking), measures the effect of the experimental variables upon its attractiveness. The third ranking, made one day later (i.e., the 24-hour

[3] Pairs of plaques, which were promised to all who volunteered to serve as *S*s, were given at the close of the second experimental session.

TABLE II

MEAN RANK OF ATTRACTIVENESS OF THE CRITICAL TASK, MADE ON NINE-POINT SCALE IMMEDIATELY AFTER AND ONE DAY FOLLOWING PERFORMANCE, UNDER EACH EXPERIMENTAL CONDITION

	Experience Success		Experience Failure	
	Expect Success	Expect Failure	Expect Success	Expect Failure
Immediate Rankings	3.6	3.3	6.4	4.3
24-hour Rankings	3.3	3.3	5.9	3.4

ranking[4]), is a measure of the permanency of the changes in the attractiveness of this task which had appeared in the immediate rankings. In both immediate and 24-hour rankings, a value smaller than five indicates that the critical task is ranked as more attractive than originally, while a value larger than five indicates that the task is less attractive. Table II gives the means of the immediate and the 24-hour rankings of the critical task for the 10 *Ss* in each of the four experimental conditions. Statistical study of the immediate and the 24-hour ranking yields the following information.[5]

General similarity of immediate and 24-hour rankings

1. The mean of all 40 rankings of the critical task has not changed significantly over the period of one day (cf. mean of immediate rankings, 4.4, with mean of 24-hour rankings, 4.0. $P > .05$).

2. The means of the immediate and the 24-hour rankings have both indicated increasing attractiveness of the critical task. The mean of the immediate rankings is 4.4, or an average +.6 increase over the original rankings of 5. The mean of the 24-hour rankings is 4.0, which is an average +.4 increase over the mean of the immediate rankings. Neither increase is in itself statistically reliable, although the increase of +1.0 between the original and the 24-hour rankings is significant ($P < .05$).

3. As was observed in the immediate rankings, experimental variation of expectation and experience is still accompanied by significant variation in the 24-hour rankings of attractiveness. (a) The critical task is still ranked as more attractive under the condition of expected failure than under the condition of expected success (cf. 3.3 and 3.4 with 3.3 and 5.9). (b) The critical task is still ranked as more attractive under the condition of experienced success than under the condition of experienced failure (cf. 3.3 and 3.3 with 5.9 and 3.4). The analysis of variance reported in Table III shows these two differences to be statistically reliable ($P < .05$).

TABLE III

ANALYSES OF VARIANCE OF ATTRACTIVENESS RANKINGS OF CRITICAL TASK MADE IMMEDIATELY AFTER AND ONE DAY FOLLOWING ITS PERFORMANCE

	Immediate Rankings		24-hour Rankings	
	F	P	F	P
Experience	11.49	<.01	5.49	<.05
Expectation	4.59	<.05	4.71	<.05
Exper. × Expec.	2.58	>.05	4.71	<.05

With 1 and 36 degrees of freedom, an F of 4.11 has a P-value of .05; an F of 7.39 has a P-value of .01.

[4] The term '24-hour' is not strictly accurate, since *S*'s and *E*'s schedules did not always permit identically timed appointments on the two consecutive days.

[5] All t-tests between immediate and 24-hour rankings have been made using the formula for correlated data.

4. Again, as in the immediate rankings, the two conditions in which expectation and experience are in contrast (one of success and one of failure) produce significantly different rankings of attractiveness (5.9 less 3.3 is +2.6, $P < .02$). This does not occur when expectation and experience are in the same direction, both of success or both of failure (3.4 less 3.3 is +0.1, $P > .05$).

Predicted similarities and differences

In the introduction, several hypotheses were advanced which related the permanence of rankings to the permanence of feelings of success or failure. These hypotheses culminated in two predictions: (1) that the rankings of attractiveness made under the condition of experienced success would not vary significantly over the period of a day; (2) that the rankings of attractiveness made under the condition of experienced failure would rise, but that the rise would be greater under the condition of expected failure, experienced failure than under the condition of expected success, experienced failure.

5. It is readily apparent that the data support Prediction 1. The rankings of attractiveness under the condition of experienced success have changed very little (cf. 3.6 with 3.3 and 3.3 with 3.3). (a) Under the condition of expected success, experienced success the mean rank has risen slightly but not significantly ($P > .05$). (b) The mean rank under the condition of expected failure, experienced success has not changed.

6. The data also support Prediction 2. The rankings under the condition of experienced failure have risen (cf. 6.4 with 5.9 and 4.3 with 3.4). But, as predicted, the rise is not uniform for (a) under the condition of expected success, experienced failure the mean rank has risen but not significantly ($P > .05$). (b) Under the condition of expected failure, experienced failure attractiveness has risen and the rise is statistically reliable ($P < .05$).

Interaction

7. With the passage of one day, three of the conditions are no longer differentiated by their mean rankings of the critical task (cf. 3.3, 3.3, 5.4, 3.4). The one strikingly different mean occurs under the condition of expected success, experienced failure, which is responsible for the continuing statistical significance of the within variable variation of expectation and experience reported above. After the period of one day it appears that pre-experimental variation of expectation no longer differentiates *Ss* under the condition of experienced success (cf. 3.3 and 3.3) but does differentiate *Ss* under the condition of experienced failure (cf. 5.4 and 3.4). This interaction between expectation and experience has attained statistical reliability ($P < .05$).

Summary and Conclusions

It was hypothesized that the permanence of experimentally induced changes in the attractiveness of a performed activity would be a function of the permanence of the feelings of success and failure originally associated with the change. Specifically, it was hypothesized that feelings of success would remain unchanged over the period of a day, and that feelings of failure would be modified, especially if failure had been expected because of the (apparent) unusual difficulty of the performed task. It was observed (a) that *Ss* under the condition of experienced success found the task equally attractive one day later,

and (b) that Ss under the condition of experienced failure found the task generally more attractive one day later; reliably more attractive if failure had been expected because of the unusual difficulty of the task.

The evidence supports the hypothesis that the permanence of experimentally induced changes in the attractiveness of activities is a function of the permanence of feelings of success and failure originally associated with the change. Since this experiment has been done under conditions conducive to limited ego-involvement, no generalization to strongly ego-involving situations has been made.

(Manuscript received for immediate publication June 21, 1949)

References

1. Gebhard, Mildred E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation and need. *J. exp. Psychol.*, 1948, **38**, 371–388.
2. Gebhard, Mildred E. Changes in the attractiveness of activities: the effect of expectation preceding performance. *J. exp. Psychol.*, 1949, **39**, 404–413.

EXPERIMENTAL STUDIES IN ROTE-LEARNING THEORY: VIII. DISTRIBUTED PRACTICE OF PAIRED ASSOCIATES WITH VARYING RATES OF PRESENTATION

BY CARL I. HOVLAND

Department of Psychology, Yale University[1]

In an earlier article in the present series (4) the writer found that paired associates benefit less from distributed practice than do serial lists. Two alternative hypotheses were advanced to explain the results. One hypothesis attributed the results to differences in the character of the two kinds of material, particularly in the extent to which they involve formation of remote associations. The alternative explanation was in terms of the rate of presentation of the units in the two procedures. In the serial learning the items were presented at a two-sec. rate. In the case of paired associates the items were also exposed at a two-sec. rate, with two-sec. exposure of the stimulus syllable and two sec. for the associated response syllable. But this meant that the subjects' *responses* were made every two sec. with serial material but only once every four sec. with the paired associates, where only the response words were said aloud. The serial lists may thus be regarded as having been presented relatively more rapidly than the paired associate lists. The writer has previously shown that more rapid presentation leads to greater economy of distributed practice (3).

As a test of the possibility that the differential advantage of distributed practice was due to the difference in rate of response, the present experiment was conducted. The rate of presentation was increased in an attempt to determine whether paired associates would benefit by distributed practice to an extent comparable to that found previously for serial lists. A one-sec. presentation rate was used for the more rapid rate, and a comparison made with a two-sec. rate, which was the speed employed in the earlier experiment.

An additional variable investigated was the rate of interpolated color-naming introduced to prevent rehearsal during distributed practice. In previous experiments in the present series the rate of color-naming was once every two sec. With this speed it may be possible for some subjects to rehearse the material silently during the interpolated activity. Speeding up the rate of presentation to once per sec. should make rehearsal more difficult and comparison of the two rates of color-naming should provide some estimate of the importance of the possibility of rehearsal.

Experimental Procedure

Paired associates of nonsense syllables were learned to a criterion of two perfect recitations with variations in the following factors:

(a) distribution of practice,
(b) rate of syllable presentation,
(c) rate of interpolated color-naming during distributed practice.

The speeds of presentation were either at a one-sec. or at a two-sec. rate. The stimulus syllables and the response syllables were given the same exposure; thus, with the one-sec. rate responses were made at two-sec. intervals, and with the two-sec. rate at four-sec. intervals. The interpolated colors during distributed practice were named at either a one-sec. or a two-sec. rate.

[1] The experimental work on which this article is based was completed in 1942. The writer expresses his appreciation to William O. Jenkins and Donald Appleton for their work as research assistants during this period.

The syllables were presented at a constant rate by means of a film projector apparatus which advanced the film frames one at a time by means of a ratchet arranged similar to that described by Schlosberg (6). The syllables from motion picture film negatives were projected on a ground glass screen cut into a wooden partition which separated the experimenter and apparatus from the subject. Each strip of film contained six different orders of the paired associates. In other respects the experimental procedure was the same as that previously described (4).

Each subject was given four days of practice to familiarize him with the general procedure. During this period he was given practice under both massed and distributed conditions at the two rates of presentation and with both slow and rapid color-naming.

In the main portion of the experiment each subject learned two paired associate lists under each of the six experimental conditions specified in Table I.

TABLE I

EXPERIMENTAL CONDITIONS

Condition	Type of Practice	Rate of Syllable Presentation	Rate of Interpolated Color-Naming During Distributed Practice
I	Massed	1 sec.	—
II	Massed	2 sec.	—
III	Distributed	1 sec.	1 sec.
IV	Distributed	2 sec.	1 sec.
V	Distributed	1 sec.	2 sec.
VI	Distributed	2 sec.	2 sec.

Each list consisted of nine pairs of nonsense syllables chosen from Melton's lists.[2] No syllables with association value above 47 percent were used. The average association value for the various lists varied from 21.8 to 29.0. The lists were rotated so that each list was employed equally often for each condition of learning.

Under massed practice conditions six sec. elapsed between successive presentations of the list. During distributed practice, a learning trial and a rest pause were regularly alternated. The rest pauses were two min. and six sec. in length. During this pause the subjects named the colors automatically presented to him.

The series of colors named during distributed practice were mounted on an electrically driven Hull memory drum which presented them at a constant speed of either one per sec. or one every two sec. The color-naming apparatus was situated directly on the subject's right so that he could name the colors without changing his position by merely turning his head slightly.

Eighteen Yale College students, paid for their services, were employed. The subjects reported at the same time each day except Sunday. They were not told the purpose of the experiment nor were they told in advance which experimental program would be employed on a given day.

[2] These unpublished materials were kindly made available in mimeographed form by Dr. Melton.

RESULTS

1. *Advantage of distributed practice with one-sec. and two-sec. speed of presentation*

For purposes of comparing the effect of distribution of practice at the two rates of syllable presentation the results for the two speeds of interpolated color-naming are combined. The comparative numbers of trials to reach the criterion of two perfect recitations by massed and distributed practice for the two speeds of syllable presentation are shown in Table II.

A sizeable difference is seen in the number of trials required to learn by massed and distributed practice at the one-sec. rate. A much smaller, but still statistically significant, difference is found between the two types of practice at the two-sec. rate. The curves of acquisition for massed and distributed practice at the two rates of presentation are presented in Fig. 1.

2. *Effect of various speeds of interpolated color-naming*

The effects upon speed of learning of two speeds of color-naming are shown in Table III. Small differences are found between the two rates of interpolated color-naming at both speeds of syllable presentation. In both cases the slower rate of interpolation results in faster learning.

Note that even with the rapid color-naming distributed practice is

TABLE II

NUMBER OF TRIALS REQUIRED TO LEARN LISTS OF PAIRED ASSOCIATES TO CRITERION OF TWO PERFECT RECITATIONS BY MASSED AND DISTRIBUTED PRACTICE AT ONE-SEC. AND TWO-SEC. RATES

	One-sec. Rate		Two-sec. Rate	
	Massed	Distributed	Massed	Distributed
Median number of trials	27.5	19	15	12
Mean number of trials	30.2	20.8	16.4	13.5
Mean difference	9.4		2.9	
S.E.$_{diff.}$	2.35		0.53	
$t_{diff.}$	4.0		5.9	
P	<.01		<.01	

superior to massed. (Mean for massed practice 30.2, for distributed 21.6, at the one-sec. rate of syllable presentation, and 16.4 for massed and 14.7 for distributed at the two-sec. syllable presentation rate.)

DISCUSSION

On the basis of the earlier results the writer inclined toward considering the difference in the *structure* of serial and paired-associate learning materials to be intimately related to the efficacy of distributed practice:

> . . . It appears that the most likely explanation of the difference in the effectiveness of distribution of practice with the two materials is the difference in the nature of the interferences involved. Long associational bonds which are prominent in serial learning have little opportunity to become formed in the present paired-associate technique. If these give rise to conflicting associations and weaken more rapidly during a short interval than do the correct associations, which are presumably stronger, distribution of practice would naturally be more effective than massed practice. (4, p. 632)

The present results incline the explanation somewhat more in the direction of the other alternative:

> The other possibility of decremental differences (after the analogy of *refractory phase*) would be based on the fact that even though paired-associates are presented at a two second rate, the response on the part of the subject only has to be made every four seconds with paired-associates, while with serial materials the response must be made at two second intervals. The more rapid rate of making responses would cause decrement in response. A close connection between the present experiment and an earlier one concerning the effect of rate of exposure upon distribution of practice would then be indicated. (4, p. 629)

The results do not, of course, indicate that serial lists and paired associates are equally susceptible to improvement through distributed practice, since there is no way of knowing what truly equivalent rates

TABLE III

NUMBER OF TRIALS REQUIRED TO LEARN PAIRED ASSOCIATES TO CRITERION OF TWO PERFECT RECITATIONS WITH TWO SPEEDS OF INTERPOLATED COLOR-NAMING DURING DISTRIBUTED PRACTICE LEARNING AT ONE-SEC. AND TWO-SEC. RATES OF SYLLABLE PRESENTATION

(A) One-sec. rate of syllable presentation

	1-Sec. Color-Naming	2-Sec. Color-Naming
Median number of trials	20	17
Mean number of trials	21.6	19.0
Mean difference	2.6	
S.E.$_{diff.}$	1.40	
$t_{diff.}$	1.8	
P	.09	

(B) Two-sec. rate of presentation

Median number of trials	13	11
Mean number of trials	14.7	12.3
Mean difference	2.4	
S.E.$_{diff.}$	0.98	
$t_{diff.}$	2.4	
P	.03	

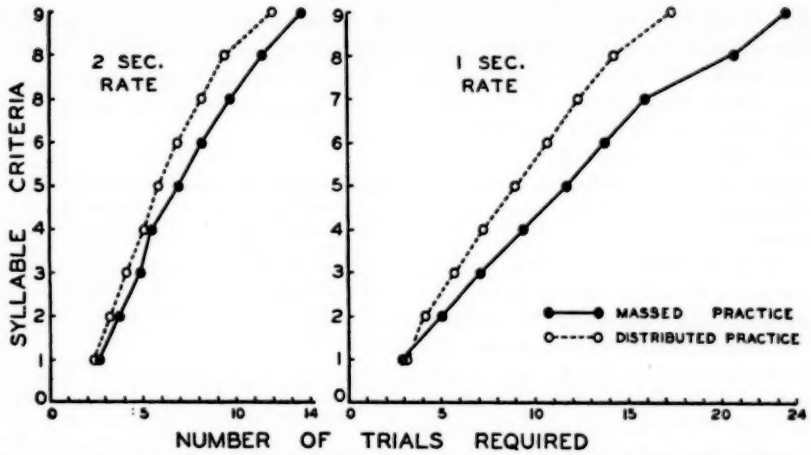


FIG. 1. Graph showing mean number of trials required to reach successive criteria of performance by massed and distributed practice with one-sec. and two-sec. rates of syllable presentation

would be for the two materials. But the results do indicate that the wide differences in the structure of the two types of learning do not produce qualitative differences in susceptibility to distributed practice, as the earlier results had suggested.

It will be observed that in the present experiment the difference between massed and distributed practice at the two-sec. rate is statistically significant. In the earlier experiment the difference was in the same direction but smaller and not statistically significant. No explanation suggests itself for the divergence in results on this point.

The lack of more sizeable differences between the groups with one-sec. and two-sec. rates of color-naming is of special interest because of the ever-present possibility that subjects can rehearse implicitly during the interpolated activity. In fact, the possibility has been investigated that the failure of Melton and Stone (5) to obtain reminiscence in their study was due to the rapid rate of color-naming which they employed (7). The investigators found, however, that speed of color-naming is not a critical factor in reminiscence. [Buxton (1) has since shown that the divergence between the results of Melton and Stone and of the writer (2) is probably attributable to the fact that the former used meaningful adjectives while the latter used nonsense syllables.] The present results corroborate those of Withey, Buxton and Elkin (7) in showing a relatively small effect attributable to speed of color-naming. Implicit rehearsal while naming colors at a one-sec. rate would be a most difficult feat. The fact that there is such a small difference in rate of learning between this condition and the two-sec. rate reduces greatly the likelihood of implicit rehearsal being an important explanation for the advantage of distributed practice in verbal learning even at the conventional two-sec. rate.

SUMMARY

1. Paired associates, which were found in a previous experiment to be

little benefitted by distributed practice when a two-sec. presentation rate was used, were learned in substantially fewer trials by distributed practice when the presentation speed was increased to a one-sec. rate. A small superiority of distributed practice was obtained even at the two-sec. exposure rate.

2. As a check on possible implicit rehearsal by the subjects during the color-naming interpolated between distributed trials, a comparison was made between learning with a one-sec. and a two-sec. rate of color-naming. A small difference favoring the slower speed was found.

3. Even with the more rapid rate of color-naming there was a pronounced advantage attributable to distribution of practice. This makes it appear unlikely that the advantage of distribution is primarily due to rehearsal during the interpolated color-naming.

4. The finding that there is not a qualitative difference between serial and paired-associate materials in their susceptibility to benefits from distributed practice makes it less likely that the presence of remote associations characteristic of serial learning is a necessary factor in distributed practice efficacy with verbal material.

(Manuscript received for immediate publication July 15, 1949)

References

1. Buxton, C. E. Repetition of two basic experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 676–682.
2. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
3. Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
4. Hovland, C. I. Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **25**, 622–633.
5. Melton, A. W., & Stone, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
6. Schlosberg, H. An inexpensive memory drum. *J. exp. Psychol.*, 1941, **29**, 161–163.
7. Withey, S., Buxton, C. E., & Elkin, A. Control of rest interval activities in experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 173–176.

THE CONSTRUCTION OF SUBJECTIVE BRIGHTNESS SCALES FROM FRACTIONATION DATA: A VALIDATION *

BY R. M. HANES

The Johns Hopkins University **

Advocates of sensory measurement claim that the fractionation procedure permits the assignment of conventional numerals to represent the magnitude of sensory experience. Others have disputed this claim, stating that such numerals do not carry conventional meaning until shown to do so. Campbell and Irwin (1), for example, in discussing the sone scale, say that if such a scale, based on guesses of $\frac{1}{2}$, represents equally well guesses of $\frac{1}{3}$ or $\frac{1}{10}$, then there would be some justification for the assumption that the '$\frac{1}{2}$' has meaning in the conventional sense. But since this is not true and because other conditions are not fulfilled, ". . . subjective measurements of loudness in numerical terms . . . are mutually inconsistent and cannot be the basis of measurement."

This criticism has not been satisfactorily invalidated. As Campbell and Irwin state, there are inconsistencies among the numerical values based on various types of estimates. Nearly all of the fractionation experiments calling for estimates other than $\frac{1}{2}$ have made use of the loudness continuum. These different sets of data have shown varying degrees of agreement between scale values from $\frac{1}{2}$ judgments and those from other fractional and multiple estimates; none has indicated more than a possibility of concurrence.

Laird, Taylor and Wille (10) obtained estimates of $\frac{1}{2}$ as loud, twice as loud, reduced by $\frac{1}{4}$ and reduced by $\frac{3}{4}$. While they found disagreement between halving and doubling, they concluded that such disagreement is not greater than would be due to experimental error. Nevertheless, if separate subjective scales are constructed graphically for each type of estimate which they obtained, serious disagreement among the four types can be shown. On the other hand, such scales based on Ham and Parkinson's (7) judgments of $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{5}$ as loud and 2, 3 and 5 times as loud apparently show excellent agreement. The degree of agreement is questionable, however, because the ranges of multiple and fractional estimates did not overlap sufficiently to permit an adequate check.

Further inconsistencies can be shown in the data from other experiments. Asking for even finer degrees of discrimination, Geiger and Firestone (6) required estimates of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and 1/100 as loud and 2, 4, 10 and 100 times as loud. Here the maximum discrepancy, as shown by scales constructed graphically, seems to occur between estimates of $\frac{1}{2}$ as loud and twice as loud. Churcher, King and Davies (3) tried successive halving and quartering of loudness, finding that scales based on the two different procedures agreed 'quite closely.'

* This research has been supported in part under the terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is Report No. 166-I-97, Project Designation No. NR 784-001, under that contract.

** The author is indebted to Dr. W. R. Garner for aid and suggestions throughout the study, to Dr. C. T. Morgan for suggestions in writing the manuscript and to Dr. A. Chapanis and Mr. Nelson Krause for suggestions concerning apparatus design.

Such is the evidence from sense-ratio data. In none of these experiments has there been adequate statistical checks on the differences among the various types of estimates. Apparently consistency has been assumed on the basis of other criteria, such as comparison with monaural-binaural, tone-combination and bisection data. Each of these procedures is, however, questionable in some degree. The monaural-binaural loudness ratio of 2:1 is an assumption. Churcher (2) has questioned the validity of using tone-combination data in obtaining loudness values. Gage (4, 5) has brought forth experimental evidence against the internal consistency of the bisection method.

Purpose

Since the $\frac{1}{2}$ estimate has come to be used almost exclusively in subjective scale construction (9, 11, 12, 13, 14, 15), the validity of the procedure needs to be investigated as thoroughly as possible. Satisfactory validation should be possible within the method itself if the numerals of the scales do carry conventional meaning. The present study was designed to determine whether or not different fractional and multiple estimates really do lead to significantly different scales.

Apparatus

Reasonable success with brightness estimates in a previous experiment (8) and a desire to extend the scale determined from that work led to the present attempt to carry out this validation for the same sensory dimension. The apparatus was designed to present a relatively simple and yet satisfactorily controlled judgmental situation. Two 150-watt projection bulbs served as the light sources. These were mounted in bases obtained from two small commercial spotlights and housed in separate metal boxes. The light from each bulb was passed successively through a sheet of opal glass, a diamond-shaped diaphragm, a second sheet of opal glass and a $1\frac{3}{4}$-in. diameter hole cut in a thin piece of sheet metal mounted against the second opal glass. These two circular openings were spaced about $4\frac{1}{2}$ in. apart at the end of a rectangular ($3\frac{1}{2}$ by 8 in.) metal box 10 in. long facing the *O*. The box was attached to a large metal bowl at the end opposite the circular openings. All metal parts were painted a flat black. The stimulus field then consisted of two luminous white discs side by side, about 22 in. from the *O*, in a dark field. Each disc subtended a retinal angle of about $4\frac{1}{2}$ degrees at the average viewing distance. The size of the diaphragm opening and consequently the brightness for each disc could be varied continuously by either *O* in front of the apparatus or by *E* at the rear. Wratten neutral density filters were provided for the lower brightness levels to allow a larger range of possible adjustment of the variable.

The brightness of the discs for the range of diaphragm openings was determined from the readings of two individuals using a Macbeth Illuminometer. At each of 22 different openings for each diaphragm an average of six readings was taken.

Procedure

Four kinds of estimates were used: $\frac{1}{3}$ and $\frac{1}{4}$ as bright and two and three times as bright. The $\frac{1}{3}$ estimate was chosen for the following purpose. In the author's previous study (8) it was found that in estimating $\frac{1}{2}$ as bright, the *O*s attempted to bisect the interval between the standard and an imaged $\frac{1}{3}$ value. This procedure is a possible explanation for the discrepancies between the scales constructed from the two types of estimate in that experiment. An odd fraction should be less likely to initiate such a procedure. Then, too, $\frac{1}{3}$ is close enough to $\frac{1}{4}$ to allow some conclusions concerning the fineness of the fractional discriminations that will be consistently brought out by a reasonable number of *O*s. The multiples were used to determine whether different operations on *O*'s part with the same numerical ratios would lead to the same scales.

Each *O* made each type of estimate but in a different order or sequence. Thus, there were 24 *O*s, each using a different permutation of order for the four kinds of estimates. Five different standard brightnesses, covering four log units, were used with each fraction and multiple. The standards for the multiple estimates were made one log unit lower than those for the fractional. Eight observations were made at each brightness level for each type of judgment by each *O*, resulting in a total of 3840 settings. The standards were presented in either an ascending or a descending order of

brightness or were mixed. Since the previous experiment had shown that this was not a significant factor, the variations were introduced primarily to prevent *O* from becoming familiar with the standards.

Prior to the fractional and multiple estimates, each *O* was required to make four equality matches by the method of average error at each of three brightness levels. This was done to give *O* familiarity with the equipment and to allow some comparison of variability between equality matches and ratio estimates. Immediately after these equality settings the fractionation procedure was started. It was suggested to *O*, as it had been with the equality matches, that he 'bracket' the value he believed to be the one called for, and he was cautioned against trying to use any criterion other than brightness in arriving at his judgments. No other instructions were given except, of course, the type of estimate required. Thus there was no possible suggestion influencing the estimates.

Every *O* had at least two experimental sessions, and some had three. In every case two observations were made at each brightness level for the first fraction or multiple called for, and then the same number was made for each of the other three types before the first one was required again. This procedure was necessary because not all of the 160 estimates for a single *O* could be made in one experimental session. Each of these series took about 45 min., on the average,

TABLE I

FRACTIONAL AND MULTIPLE ESTIMATES AND EQUALITY MATCHES

Each average is based on 24 individual means of 8 observations each.

(All values are in millilamberts.)

Type of Estimate	Statistic	Standard Brightness				
		0.05	0.5	5	50	500
One-half	Mdn.	0.013	0.120	0.684	5.556	66.72
	M	0.015	0.118	0.795	6.988	80.50
	σ	0.006	0.044	0.542	4.132	48.16
	σ_m	0.001	0.009	0.113	0.862	10.04
One-third	Mdn.	0.009	0.060	0.313	2.375	18.34
	M	0.009	0.068	0.390	3.160	34.43
	σ	0.004	0.038	0.339	2.398	41.78
	σ_m	0.001	0.008	0.071	0.500	8.71
		Standard Brightness				
		0.005	0.05	0.5	5	50
Double	Mdn.	0.042	0.290	3.286	47.24	281.0
	M	0.056	0.300	3.339	53.13	279.9
	σ	0.036	0.096	1.392	30.93	95.6
	σ_m	0.007	0.020	0.290	6.45	19.9
Triple	Mdn.	0.118	0.579	6.80	136.6	493.6
	M	0.168	0.856	10.54	136.8	464.8
	σ	0.157	0.621	9.13	74.2	148.4
	σ_m	0.033	0.130	1.90	15.5	30.9
		Standard Brightness				
		5		50	500	
Equality Matches	Mdn.	5.14		48.54	461.6	
	M	5.28		50.02	472.2	
	σ	0.68		5.89	48.6	
	σ_m	0.14		1.23	10.1	

with a few *O*s taking as few as 25 min. and others needing over an hour for the same number of readings. No time limit of any kind was imposed; each *O* was instructed to take as long as he needed.

The standard was shifted between right and left on successive series of brightness levels. The apparatus was turned off between judgments and the setting of the variable was randomly altered before each succeeding presentation. While all observations were made in a dark room, no attempt was made to maintain any given level of adaptation since this would be practically impossible with the procedure used here.

The whole design was such as to provide the barest minimum in cues for consistency other than the setting based on the standard brightness presented each time; it is obvious from the results that some measure of success was obtained.

Observers

All of the *O*s were male graduates or undergraduates between 18 and 26 years of age. Only 7 of the 24 had ever served in a psychological experiment before and then only on one or two experiments unrelated to the present study. All had normal visual acuity or were corrected for defects: all reported normal color vision except one who had never been tested.

Results

Estimates in physical units.—The principal results are summarized in Table I, where measures of central tendency and dispersion are listed for the group of 24 *O*s. Some idea of the great inter-individual variability can be obtained from the size of the standard deviations. The mean value at each brightness level for any one type of estimate is, nevertheless, significantly different from the means at every other brightness level. At each brightness level, also, the difference between the means for $\frac{1}{2}$ and $\frac{1}{3}$ judgments and for double and triple judgments is significant. All these differences are significant beyond the one percent level of confidence. With these absolute values converted to ratios (between the standard bright-

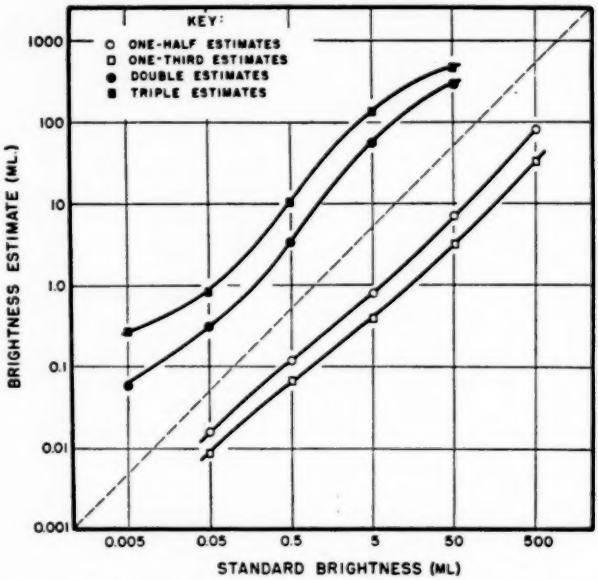


Fig. 1. Fractional and multiple brightness estimates. Each point represents the mean of 24 individual means of 8 estimates each. The dotted line indicates equal brightness.

ness and the fractional or multiple estimate) an analysis of variance showed further that the sequence or order in which the judgments were made was also a significant factor. This fact may be due to interaction variance between *O*s and type of estimate, since this variance cannot be separated from that for sequence.

With the ratio values separated into two parts, one for the fractional and the other for the multiple estimates, the variances for other factors were tested. All variances were found to be significant, i.e., those for *O*s, type of estimate, brightness level and the various interactions. The kind of response called for in the fractionation procedure is obviously easily influenced. And yet, as further analysis of the results shows, on the average the numerical values assigned have consistency and meaning.

The mean values from Table I are shown graphically in Fig. 1. From the figure it appears that the curves for the multiple values show the same trends of curvature as do those for the fractional vaues, but the trends are considerably accentuated in the former. Agreement between the estimates of ½ as bright and twice as bright (or between ⅓ and three times as bright) is best at the middle and higher brightness values. There is the tendency for nearly all multiple estimates to give higher brightness ratios (between the standard and estimated brightnesses) than do the fractional ones. It is only at the lowest brightnesses, on the other hand, that this tendency becomes so marked as to present a serious problem. On the basis of a few simple assumptions, a ready explanation for this discrepancy can be found in terms of rod and cone function.

If it is assumed that *O* looks primarily at the variable stimulus while he is making an adjustment, then the lowest brightness standard should have an enhanced relative brightness due to the fact that its image will fall somewhat more toward the periphery of the retina and will therefore, presumably, stimulate a larger proportion of rods than will the variable. Because of this, when a multiple estimate is being made, a relatively higher brightness will be required for this multiple. Such a situation would result in a ratio of physical brightnesses which would be greater than that found if the effect were not present. The same reasoning would lead one to expect relatively smaller ratios when a fractional estimate is made. In the latter case, however, the effect should not be so marked, if it is present at all, since the brightnesses involved are not generally as low as those for the multiples. This fact can be readily seen in Table I.

It might be expected that the reverse effect would be found with the higher brightness values, since the standard could presumably stimulate a smaller proportion of cones and therefore have a reduced subjective brightness relative to the variable. There is some evidence for this in the differences between the fractions and multiples, but there is much more conclusive evidence from the equality matches. For example, the mean value equated to 500 millilamberts was 472 millilamberts. There are factors, then, at least adding to, even if not wholly accounting for, the discrepancies between the fractional and multiple judgments, and they are factors apart from any inconsistencies in meaning between the fractions and multiples.

Interesting evidence of a satisfying degree of consistency for ½ estimates, at least, is shown in Fig. 2, where the ½-judgment function of the previous

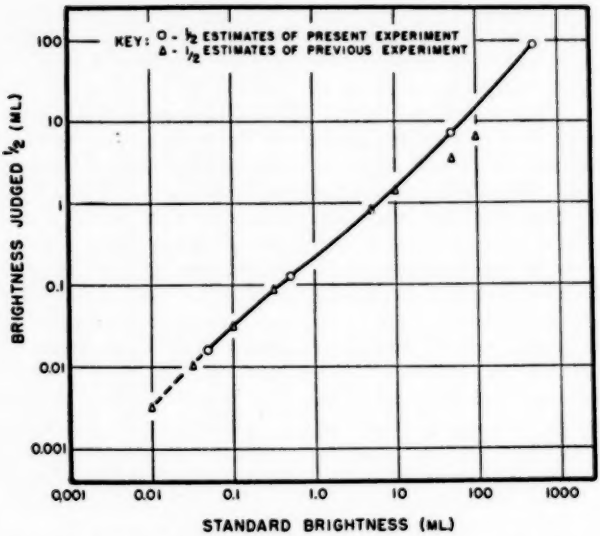


FIG. 2. Comparisons of ½ estimates from present and previous data

experiment (8) is plotted for comparison with that of the present experiment. The conditions of the previous study were quite different from those of the present one. In the former, monocular, successive comparisons were used with a 3½ degree stimulus on a cathode-ray tube. In addition, *E* made the adjustments at the direction of *O*. Four *O*s were used, and each made twenty estimates at each brightness level. The amount of agreement manifest in the two cases is rather potent evidence in favor of a stable ½ estimate of brightness under quite dissimilar conditions.

Estimates in subjective units.—So far the treatment of the results involves nothing more than describing and testing differences among various types of brightness estimates. Since the primary purpose of obtaining such data is, rather, the assignment of numerical values to represent sensory experience, the fractional- and multiple-judgment functions were used to construct subjective scales by the graphical method described by Stevens, Volkmann, and Newman (14) and Reese (12). Because this procedure has been adequately described before, it will not be reviewed here. It should be pointed out, nevertheless, that, being graphical, the procedure is limited in accuracy and the method of obtaining the scale numbers is such that errors tend to multiply.

Individual scales for each type of judgment were constructed for each *O*. This procedure resulted in a total of 96 such scales, 24 each for ½, ⅓, double and triple estimates. The purpose of this procedure was to obtain a distribution of numerical values assigned to each of several selected physical brightness values, in order to determine whether the different types of estimates would result in significantly different scale values. Each scale was anchored at the same point by arbitrarily setting

0.05 millilamberts (the lowest convenient value) equal to the number 25, i.e., 25 brils to use the name suggested previously for the subjective units. Anchoring the scales at this point leads to maximal discrepancies among the numerals assigned to the points tested, since the greater the distance from an anchor point, the greater the possible divergence. In other words, anchoring all scales nearer the middle of the physical brightness range would tend to minimize dissimilarities.

Two distributions of 96 scores each were obtained from these scales, one at 100 millilamberts, the other at 200 millilamberts. The analysis of variance showed, in both cases, highly significant differences among *O*s but no significant differences due to type of estimate. That is to say, fractional estimates of either $\frac{1}{2}$ or $\frac{1}{3}$ as bright and multiple estimates of either two or three times as bright result in scales which are not, on the average for a number of *O*s, significantly different. Individual's scales are, on the other hand, significantly different from one another.

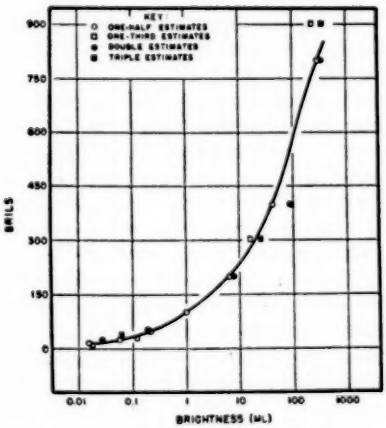


FIG. 3. Subjective scales from the various fractional and multiple estimates

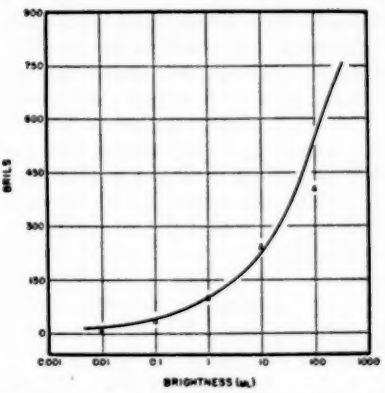


FIG. 4. Bril scale from the combined data for the four types of estimates. Points from the previous scale (triangles) are plotted for comparison.

Fig. 3 shows the agreement, or lack thereof, among the mean scales constructed from the four kinds of judgments. These scales are anchored at one millilambert equal to 100 subjective units in order to allow more convenient comparison with the scale previously mentioned (8). This comparison is presented in Fig. 4. The solid line in Fig. 4 is the average of the four scales indicated in Fig. 3. It is evident that the two different experiments resulted in essentially the same scale up to about 50 millilamberts. There is no obvious explanation for the lack of agreement beyond that point, but since the former scale was based on the results of only four *O*s, the present scale is considered to be the more reliable.

DISCUSSION

After the data had been collected for an *O* he was asked a series of questions for the purpose of determining, if possible, how the estimates were made. All but two of the *O*s declared that they had heeded the warning in the instructions and, so

far as they knew, had used only brightness as a criterion for their settings. These two exceptions admitted that a few times they had tried mechanical (number of turns of the adjusting knob) means but that this procedure was not satisfactory. As mentioned earlier it was expected that no one would try to image ½ in setting ⅓. This was found to be the case, but about half of the *O*s either always or sometimes first adjusted to ½ when making a ⅓ judgment and/or adjusted to twice as bright when making a triple judgment. No apparent differences are to be found in comparing the results from these *O*s with the data from those who did not follow this procedure.

It might seem at first thought that estimating a ratio of 2:1 would be the same whether fractional or multiple estimates are called for, since the *O* is free to look back and forth in this situation and think either in terms of the variable, in the case of multiple estimates, for example, as being the required multiple of the standard or in terms of the standard being the required fraction of the variable. It was hoped that such 'cross comparisons' would not occur, in which case the multiples and fractions would be more distinctly different operationally. It is interesting to note that only nine of the *O*s ever used such comparisons. In fact, a few expressed surprise and even some chagrin at not having thought of such a procedure. Of these nine cases, only two used these comparisons regularly for both fractional and multiple settings; three others did the same some of the time. The remaining four regularly thought in terms of multiples when making fractional extimates but not the reverse. This would seem to indicate that, if anything, the multiple estimates are considered somewhat easier by the *O*s. This is borne out to some extent by verbal reports, slightly more of the *O*s preferring the multiples as being easier. There is also a tendency for the type of estimate (fractional or multiple) called for first to be considered less difficult, and the ratio 3:1 is pretty generally judged more difficult than 2:1.

Seventeen of the *O*s made a hurried equality match between the standard and the variable each time or most of the time before proceeding to the required setting. Extremely interesting is the fact that the standard deviations around the group means of the equality matches required of each *O* at the start (5, 50, and 500 millilamberts) are of the same order of magnitude as the deviations around the means of the ½ and ⅓ estimates at the same brightness levels. On the basis of this fact it would appear that the equation of brightness in the present situation, at least at relatively high levels, is nearly as variable as are fractional judgments. Perhaps there is not as much difference as has often been claimed between the two types of responses.

The procedure of making equality matches first might suggest judging in terms of addition or subtraction, but the implication is not altogether clear. Two subjects did report thinking in terms of adding brightnesses; another stated that he attempted to imagine the two discs superimposed but that it 'didn't work.' None of the *O*s was very clear as to just how he did make his estimates. Since agreement has been shown among the scales from the various types, the mechanism must be quite subconscious.

Expressions of confidence (as to the agreement of their settings with actual physical values) by the *O*s ranged from plus or minus 10 to plus

or minus 1000 percent, with little correlation between expressed degree of confidence and accuracy of estimate. One very interesting *O*, on the contrary, did express considerable confidence in his judgments, and he was right, his estimates being remarkably close to the physical fractions and multiples. Since this *O* was so atypical, his data naturally have a great effect on the variability of the group, especially in the scale values obtained. This *O* reported that he had been a Navy signalman for a year and therefore had considerable experience with lights at various distances. So far as he was able to determine, he had not used such experience in the present situation, but the very fact that he thought of the comparison is suggestive. The data for this single *O* can account for part of the discrepancy shown in Fig. 2 between the present results and those of the previous experiment, since his ½ estimates were considerably higher than those of any other subject, especially at the higher brightness values.

Considering the inexperience of the *O*s and the minimal instructions given, this experiment can certainly be considered as a reasonably exacting test for the validity of estimating sense ratios, and it must be admitted that the technique has not escaped entirely unscathed. Individual differences have been found, for example, to be considerably greater than might have been expected from the previous experiment dealing with brightness estimates. The sense-ratio judgment is obviously quite variable and subject to a host of influences. And yet, on the average, the results are remarkably consistent. Further work is called for before the procedure can be finally accepted. It would be interesting, for example, to apply the procedure used here to estimates of pitch, which might not be subject to quite so many influences as are the brightness estimates.

Summary and Conclusions

An attempt has been made to validate the fractionation technique for the construction of sensory scales by statistical comparison of the results obtained from estimates of ½ and ⅓ as bright and two and three times as bright. The results for 24 *O*s, each of whom made each type of estimate, were subjected to an analysis of variance. Since the differences between the ½ and ⅓ estimates and between the double and triple estimates were found to be significant, the construction of scales based on these different ratios was possible. In addition, the order in which the different types of estimates were made, the brightness level, individual differences and the various interactions of these factors were found to be significant.

The individual subjective scale values, obtained for each type of estimate by each *O*, were also treated by the analysis of variance. The variance due to individuals was again significant, but the variance for type of estimate was not significant. On the basis of these findings it is to be concluded that:

1. The estimation of fractional and multiple brightness is feasible and leads to consistent results for group means.

2. The estimates are, nevertheless, very easily influenced by a large number of factors.

3. Individual differences are significant and in some cases may be quite large.

4. Even with this great variability, group estimates of ½ are significantly different from estimates of ⅓, and

double estimates are significantly different from triple estimates.

5. The various fractional and multiple estimates used in the present experiment lead to subjective scales which are not significantly different. Therefore, the results are considered as a validation of the fractionation procedure for the construction of scales to represent sensory experience.

6. The subjective scale obtained from the present results is in essential agreement with the previous scale (8) and extends the range of physical brightness covered to an upper limit of about 500 millilamberts.

(Manuscript received for immediate publication July 11, 1949)

References

1. Campbell, N. R., & Irwin, J. O. On the sone scale. Appendix I. Pp. 337–338. In: Quantitative estimates of sensory events. Final report. *Advance Sci.*, 1940, No. 2, 331–349.
2. Churcher, B. G. A loudness scale for industrial noise measurements. *J. acoust. Soc. Amer.*, 1935, 6, 216–226.
3. Churcher, B. G., King, A. J., & Davies, H. The measurement of noise with special reference to engineering noise problems. *J. Inst. elec. Eng.*, 1934, 75 (454), 401–422.
4. Gage, F. H. An experimental investigation of the measurability of auditory sensation. *Proc. Roy. Soc.*, B, 1935, 116, 103–122.
5. Gage, F. H. An experimental investigation of the measurability of visual sensation. *Proc. Roy. Soc.*, B, 1935, 116, 123–138.
6. Geiger, P. H., & Firestone, F. A. The estimation of fractional loudness. *J. acoust. Soc. Amer.*, 1933, 5, 25–30.
7. Ham, L. B., & Parkinson, J. S. Loudness and intensity relations. *J. acoust. Soc. Amer.*, 1932, 3, 511–534.
8. Hanes, R. M. A scale of subjective brightness. *J. exp. Psychol.*, 1949, 39, 438–452.
9. Harper, R. S., & Stevens, S. S. A psychological scale of weight and a formula for its derivation. *Amer. J. Psychol.*, 1948, 61, 343–351.
10. Laird, D. A., Taylor, E., & Wille, H. H. The apparent reduction of loudness. *J. acoust. Soc. Amer.*, 1932, 3, 393–401.
11. Lewis, D. R. Psychological scales of taste. *J. Psychol.*, 1948, 26, 437–466.
12. Reese, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.*, 1943, 55, No. 251.
13. Stevens, S. S., & Volkmann, J. The relation of pitch to frequency: A revised scale. *Amer. J. Psychol.*, 1940, 53, 329–353.
14. Stevens, S. S., Volkmann, J., & Newman, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, 8, 185–190.
15. Taves, E. H. Two mechanisms for the perception of visual numerousness. *Arch. Psychol.*, 1941, 37, No. 265.

THE EFFECT OF PSYCHOPHYSICAL METHOD AND CONTEXT ON PITCH AND LOUDNESS FUNCTIONS [1]

BY J. M. DOUGHTY [2]

Franklin and Marshall College

Introduction

In a recent experiment (1) on the effect of duration on the pitch of a tone it was found that the method used made a difference in the results obtained. With the Method of Constant Stimuli the Constant Error (CE) was small throughout the range of durations used, increasing very slightly at the short durations. With the Method of Average Error, the CE was of much greater magnitude at the crucial short durations.

This difference posed the problem of selecting those data really representing the relation between pitch and tonal duration. The selection of the Average Error data was made on the basis of two considerations: (1) the greater relevance of the *O*'s task to the determination of points of subjective equality (PSE); and (2) the possibility that an artifact was present in the experiment in which the Method of Constant Stimuli was used.

With regard to this last point, in the Method of Constant Stimuli it is customary for the *E* to select in advance comparison stimuli which are distributed symmetrically about the standard in physical terms. This means that the total context of stimuli (which includes all standard and comparison stimuli) comprising any given experimental run is balanced in terms of frequency or intensity about the standard frequency or intensity. Unfortunately, physical symmetry does not necessarily mean psychological symmetry. This could be the case when there is some change in quality (pitch, for example) as a function of the independent variable (duration, for instance). Of course, the *E* cannot compensate for this psychological asymmetry in the stimulus context because he does not know the direction and magnitude of the change in quality occurring as a function of the independent variable. That is what he is trying to find out. That such psychological asymmetry biases the results is suggested by the above experiment and by other material that can be seen in the literature on psychophysics.

Interaction Effects

The background to the problem of the effect of context on the psychophysical judgment is to be found in the literature on interaction effects. Such effects refer to the influence upon any judgment in the experiment exerted by any one or more stimuli occurring prior to the making of the judgment. One may divide these effects into (1) intra-serial effects —or those interactions taking place between stimuli in a given series; and (2) interserial effects—or those taking place between one series of stimuli and another. Fundamentally, these are both alike in that they act as though they establish a standard for the *O* which is not necessarily the same as the experimental standard.

There is, of course, the possibility of extraserial effects taking place. That is, some general level of reference which the *O* brings into the laboratory with him affecting his judgments. It is difficult to find evidence for such effects in the typical psychophysical experiment, however, because of the rapid appearance of inter- and intraserial effects.

[1] This research has been supported in part under the terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is Report No. 166-I-100, Project Designation No. NR-784-001, under that contract. This paper is a thesis submitted in partial fulfilment of the requirements for the PhD degree at The Johns Hopkins University.

[2] The writer wishes to express his appreciation to Dr. W. R. Garner for his assistance in formulating the problem and seeing it to a conclusion.

Interserial effects.—The evidence for the presence of these effects is quite conclusive. We need not attempt a complete review but simply refer to several studies which illustrate the effect. Needham (7), and Koester (4) found that the CE for auditory intensity varied with intensity level and Woodrow (11) demonstrated the same effect in judgments of weight. In the analysis of the *O*'s judgments in such experiments as these, it was found that low levels of stimulation preceded by a series of judgments at high levels were accompanied by an increase in the number of 'softer' (or 'lighter') judgments as compared to the situation in which the low level series had not been preceded by higher levels. Conversely, the effect of prior low level stimulation upon judgments at a higher level was to increase the number of 'louder' (or 'heavier') judgments.

Intraserial effects.—This kind of effect is more relevant to the discussion of the effect of asymmetry in the stimulus context introduced in the opening paragraphs of this paper. Once again, there is quite conclusive evidence for the presence of such interaction effects in psychophysics. Rose and Harris (9) recently reviewed this material.

By way of illustration, Koester and Schoenfeld (6) demonstrated that pitch judgments were affected by preceding levels of stimulation in the same experimental series. They used several standards in a given run instead of the usual one. The judgments at a low standard level were preponderantly 'lower' following stimulation at a high standard level. Conversely, the judgments at a high standard level were preponderantly 'higher' following stimulation at a low standard level. In one more example, Fernberger (2) had arranged weights so that the comparison stimuli which ordinarily would be undecided followed several either very heavy or very light stimuli. The ordinarily undecided judgments were preponderantly 'heavier' following stimulation by the light weights, and preponderantly 'lighter' following stimulation by the heavy weights.

The Effect of the Psychophysical Method

The results of studies making comparisons betweeen psychophysical methods indicate that the method which one uses in attacking a given problem can markedly influence the results. On the face of it, one might expect the classical methods to give essentially the same data.

Postman (8), for instance, found that the usual negatively directed p-function for loudness as obtained by the Method of Constant Stimuli was absent when the method of Average Error was used. In the latter case, there appeared to be no change in the CE with time. Doughty and Garner (1) found that CEs for pitch as a function of tonal duration were all consistently smaller for the Method of Constant Stimuli in contrast to the Method of Average Error. Stevens (10), Koester and Schoenfeld (5), and Gault and Goodfellow (3)—to name a few more studies—all provide additional evidence on the point so that there is not much doubt that the method used is an important determiner of the results which one obtains in a psychophysical experiment.

Purpose

Our experience with the Method of Constant Stimuli in determining the effect of tonal duration on pitch (1) led us to the conclusion that this method should not be used for determining CEs. The basis for this conclusion rests on the known effects of context on psychophysical judgments and the arbitrary selection of the context by the *E* which occurs in the Method of Constant Stimuli.

In the ordinary use of this method, the *E* selects the context for the *O* by arranging the comparison stimuli so that they are symmetrical about the standard in physical terms. If there is a CE present in these comparison stimuli, however, they are not sym-

metrical about the standard insofar as the experience of the *O* is concerned. In view of the kind of evidence that exists on context effects, we would expect that such a psychologically unbalanced context would seriously influence the *O*'s judgments. In such a case, the experimentally determined CE would not be the true CE.

On the other hand, it was felt that the Method of Average Error does not suffer from such predetermination of the CE. Typically, in this method the *O* has complete control over the variable stimulus. The *O*, therefore, and not the *N* determines the context in which the judgments are made. In general, the *O* will experience as many stimuli higher than the standard as lower than the standard—higher or lower with respect to the quality in question since that is the only cue that the *O* has. For this reason, we would expect that the *O* makes his judgments in a balanced context insofar as his experience is concerned.

We predicted that the *E* can predetermine the CE which he will obtain by employing the Method of Constant Stimuli. That is, if the *E* selects the comparison stimuli so that they are preponderantly higher in some quality than the standard, the CE will tend to be more positive in proportion to the degree of preponderance of high stimuli. Conversely, if the comparison stimuli are mostly low the CE will be negative in proportion to the degree of preponderance of low stimuli.

This experiment was designed primarily as a test of this prediction. Consequently, we proposed to determine the CEs for pitch and loudness under varying conditions of asymmetry in the stimulus context. We were interested in comparing these context effects in the two psychophysical methods of Constant Stimuli and Average Error, under varying conditions of spread in the comparison stimuli, and for two stimulus durations.

Description of the Experiment

Apparatus

Fig. 1 is a block diagram of the apparatus used in this experiment. Two oscillators were used to generate the tones in the pitch runs. A Hewlett-Packard audio oscillator generated the standard tone and a General Radio beat-frequency oscillator generated the comparison tone. For the loudness runs, only the latter oscillator was used. The output of the oscillators (or oscillator) was led into the two channels of an electronic audio timer which allowed for the control of the duration, separation, and repetition of the stimuli. From the timer the tones were conducted separately to two 0.1-db step attenuators, then to a mixer attenuator, and finally into a speaker (for monitoring) and a Permoflux earphone (type PDR-8) for the *O* in a sound-deadened room. Connected in parallel with the earphones for the purpose of calibration was a Ballantine Voltmeter and Dumont (Type 208-B) cathode-ray oscillograph.

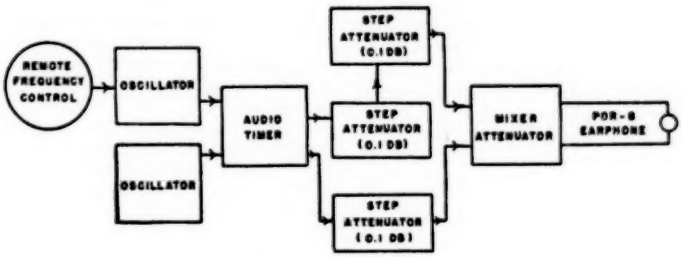


Fig. 1. Block diagram of the apparatus

For the Method of Average Error runs, a remote control of frequency and an additional attenuator were located in the sound-deadened room for the use of the *O* in making pitch and loudness matches.

Procedure

Method of constant stimuli.—The *O* was seated in the sound-deadened room and asked to adjust the head-phone so that it fitted comfortably. He was instructed to use the same ear at all times. He was told to listen to the standard stimulus and to judge whether the comparison tone was higher or lower in pitch (or louder or softer in loudness). He was told to record this judgment in the appropriate place on the tally sheet.

In a single run the *O* made seventy such judgments—10 judgments per point at 7 points. In all, there were five runs for each *O* for each set of conditions, making a total of 50 judgments per point per *O*.

Method of average error.—The procedure was the same as that for the Method of Constant Stimuli except, of course, the *O* adjusted the variable stimulus until the two tones were matched in pitch (or loudness). One judgment was made at a time and a new set of conditions were set in before the *O* was asked to make another judgment. Between judgments he removed the earphones. Sets of 12 such judgments were regularly alternated with the Method of Constant Stimuli runs. Each *O* made a total of 50 judgments per point.

The *E* used the fine vernier adjustment on the beat-frequency oscillator in combination with the oscillograph in determining the *O*'s setting for his pitch matches. These were recorded in terms of the frequency to match the standard. In the case of loudness, the *E* took readings from the Ballantine Voltmeter, recording these in terms of intensity level to match.

Experimental Variables

Degree of asymmetry: Method of Constant Stimuli.—This refers to the degree of skewness in the distribution of the comparison stimuli about the standard. There were seven degrees of asymmetry defined as follows in terms of the number of comparison stimuli higher than, equal to, and lower than the standard stimulus:

Degree	Higher	Equal	Lower
0	3	1	3
−1	2	1	4
−2	1	1	5
−3	0	1	6
1	4	1	2
2	5	1	1
3	6	1	0

Degree of asymmetry: Method of Average Error.—For this method, there were three degrees of asymmetry in the stimulus context, defined as follows:

(0)—The *O* was started at a high or low point and was instructed to bracket the standard as symmetrically as possible in making the match.

(−)—The *O* was started at a point lower than the standard and was instructed to change the variable in one direction only until the match was made.

(+)—The *O* was started at a point higher than the standard and was instructed to change the variable in one direction only until the match was made.

Comparison stimulus durations. Comparison stimulus durations of 50 msec. and 1000 msec. were used to provide two levels of difficulty in making judgments.

Degree of spread: Method of Constant Stimuli.—This refers to the number of cycles or the number of decibles over which the variable stimuli are spread, in equal steps. There were two degrees of spread:

Narrow—The limits of this range were defined as the highest and lowest points at which the *O* could just make 100% accurate discriminations. The five intermediate comparison stimuli were spread over this range in equal steps.

Broad—The narrow range was doubled for each *O*.

Degree of spread: Method of Average Error.—For pitch, the comparison stimulus was variable over a total range of 100 cycles for the narrow condition and over a range of 400 cycles for the broad condition. For loudness, the narrow range was about 10 db and the broad range about 20 db.

Comparison stimuli: Method of Constant Stimuli.—The following values are for the narrow range of stimuli. (Double them for the broad range.)

For the 1000-msec. tone, the comparison stimuli for the loudness runs were spread over 1.8 db in 0.3-db steps. The comparison stimuli for pitch were spread over 6 cycles in 1-cycle steps for two *O*s and over 12 cycles in 2-cycle steps for one *O*.

For the 50-msec. tone, the comparison stimuli for the loudness runs were spread over 4.2 db in 0.7-db steps. The comparison stimuli for pitch were spread over 15 cycles in 2.5-cycle steps for two *O*s and over 18 cycles in 3-cycle steps for one *O*.

Experimental constants.—For all judgments the following were constant:

The duration of the standard tone at 1000 msec.

The frequency of the standard tone at 1000 cps.

The intensity of the standard tone at 80 db re 0.0002 dyne/cm^2.

The interval between cessation of the standard tone and beginning of the comparison tone, at 1000 msec.

The repetition rate at 10 per min.

The mode of listening as monaural.

Training procedure.—Three *O*s were used throughout the experiment. They were all males. One was a high school student and the other two were college students. All three showed normal audiograms.

Each *O* was given a period of training in making pitch and loudness judgments. At the same time, they were gaining familiarity with the procedures for the two experimental methods we used. This included experience with both comparison tone durations and with varying degrees of asymmetry. This period ended and the actual experiment began for each *O* when he had reached the point of maximum discrimination, as defined by no further improvement.

During their training, the *O*s were told that it was quite possible for some of the Method of Constant Stimuli runs to have a preponderance of stimuli higher or lower than the standard. They were repeatedly instructed not to be concerned with the distribution of their judgments (the number of 'highs' or 'lows,' for instance) but to make all their judgments only with respect to the relation between each standard-comparison pair as they came along.

In the training period for the Method of Average Error, and throughout the experiment, the *O*s were told that their primary task was to make the equation in pitch or loudness. This task was to take precedence over moving the variable stimulus in one direction only when they were judging under the conditions of asymmetry outlined above for this method.

Analysis of data.—For the Method of Average Error data, CE's were computed from the mean of the distribution of judgments for each of the *O*s. The DL's were taken as the PE of the distribution of each *O*'s judgments. An average of the individual CE's and DL's was computed for the presentation of the results in graphical form.

The least squares solution for the phi-gamma function, with unweighted values, was used to determine the CEs and DLs. The CE was taken as the 50 percent point and the DL as one-half the interval between the 25 and 75 percent points. Each *O*'s data were analyzed in this way and an average of the individual CE's and DL's was computed for graphical presentation.

Results

Mode of Presentation

Method of Constant Stimuli.—The effect of asymmetry in the stimulus context as it occurs in this method is presented in two ways.

1. The effect of context on the distribution of 'higher' and 'lower' (or 'louder' and 'softer') judgments is shown in Figs. 2 and 3. The percent of 'higher' (or 'louder') judgments is presented as a function of the degree of asymmetry. For purposes of clarity, the percentage of stimuli in the context which are greater (i.e., of higher frequency or more intensity) than the standard is given for each degree of asymmetry. (One-half of those comparison stimuli which were equal to the standard were counted as greater than the standard.)

If the context has no influence at all on the *O*'s judgments, the slopes of the curves in Figs. 2 and 3 should be equal to 1.00. That is to say, the percent of 'higher' (or 'louder') judgments should be equal to the percent of stimuli in the context greater than the standard.

If the context influences the judgments maximally, the curves should have a slope of 0.00. In other words, the same percent of 'higher' (or 'louder') judgments should be called regardless of the percent of stimuli in the context which are greater than the standard.

2. The effect of context on the size and sign of the CEs is displayed in Figs. 4 and 5. These graphs show the relations between the CE and the arithmetic mean of the comparison stimulus series. In other words, they show the relation between the CE and the degree of physical displacement of the context from physical symmetry. To make comparisons simpler, the arithmetic mean is expressed as the

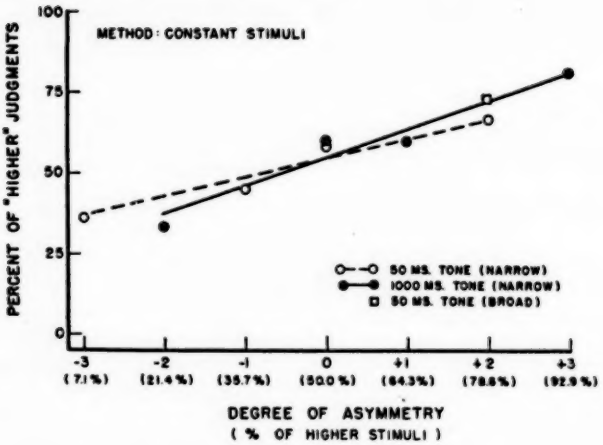


FIG. 2. The percent of 'higher' judgments of pitch as a function of asymmetry. This relation for a 50-msec. and 1000-msec. comparison tone is shown by the broken and solid lines respectively. These points were obtained with the comparison stimuli spread over a relatively narrow range. The unfilled square represents the point obtained for a 50-msec. comparison tone with the comparison stimuli spread over a relatively broad range.

difference between it and the standard stimulus.

In this case, if there is no effect of context, the slopes of the curves in Figs. 4 and 5 should be 0.00. This is to say that the CE should remain the same under all conditions of asymmetry.

If the effect of context is maximal, the slopes should be equal to 1.00. This would mean that the CE would always equal the difference between

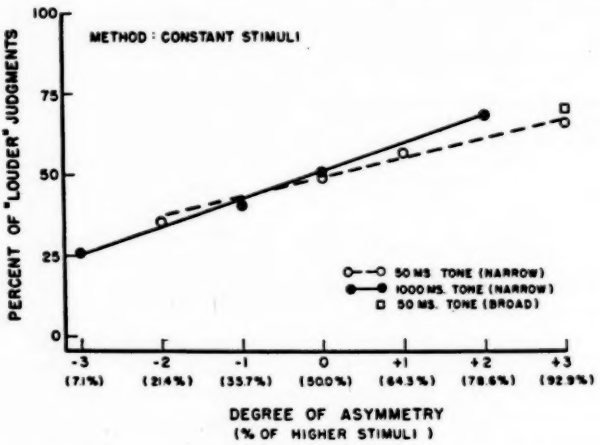


FIG. 3. The percent of 'louder' judgments of loudness as a function of asymmetry. The parameters are the same as those in Fig. 2.

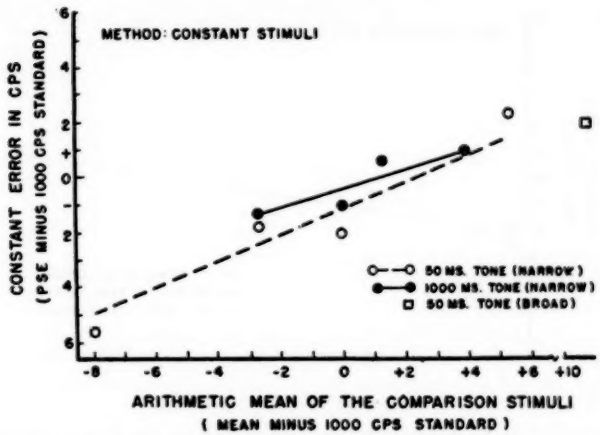

METHOD: CONSTANT STIMULI
CONSTANT ERROR IN CPS
(PSE MINUS 1000 CPS STANDARD)
50 MS. TONE (NARROW)
1000 MS. TONE (NARROW)
50 MS. TONE (BROAD)
ARITHMETIC MEAN OF THE COMPARISON STIMULI
(MEAN MINUS 1000 CPS STANDARD)

FIG. 4. The CE for pitch as determined by the Method of Constant Stimuli as a function of the arithmetic mean of the comparison stimulus series. The parameters are the same as those in Fig. 2.

the standard and the arithmetic mean of the context.

Method of Average Error.—The data from this method cannot be presented in any exactly comparable manner. There is no equivalent measure of the proportions of 'higher' (or 'louder') judgments nor is there an exact measure of the mean of the physical context. We have displayed these data simply in the form of the relation between the CE and the degree of asymmetry (Figs. 6 and 7). The degrees of asymmetry refer to the conditions in which the *O* was permitted to move the variable in one direction only starting from a point higher than the standard—(+) Asymmetry; or lower

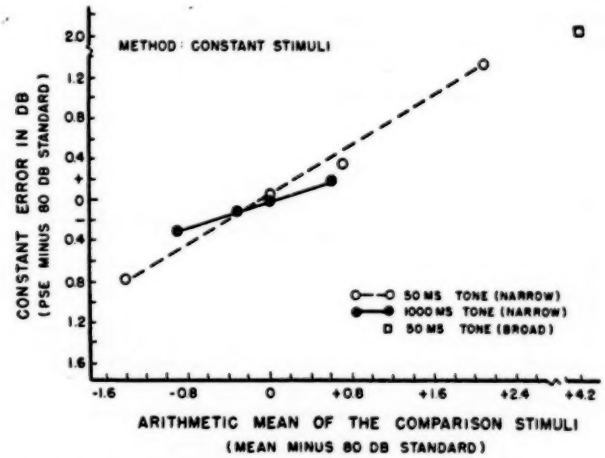

METHOD: CONSTANT STIMULI
CONSTANT ERROR IN DB
(PSE MINUS 80 DB STANDARD)
50 MS TONE (NARROW)
1000 MS TONE (NARROW)
50 MS TONE (BROAD)
ARITHMETIC MEAN OF THE COMPARISON STIMULI
(MEAN MINUS 80 DB STANDARD)

FIG. 5. The CE for loudness as determined by the Method of Constant Stimuli as a function of the arithmetic mean of the comparison stimulus series. The parameters are the same as those in Fig. 2.

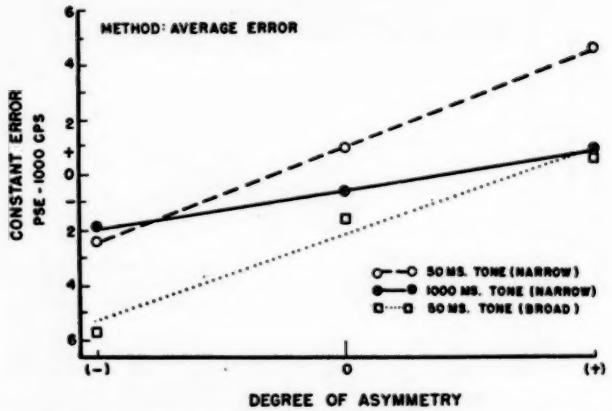


FIG. 6. The CE for pitch as determined by the Method of Average Error as a function of asymmetry. This relation for a 50-msec. and 1000-msec. comparison tone is shown by the broken and solid lines respectively. These points were obtained with the comparison stimuli variable over a relatively narrow range. The dotted line represents this relation for a 50-msec. comparison tone with the comparison stimuli variable over a relatively broad range.

than the standard—(−) Asymmetry; and the condition in which he was instructed to move the variable alternately higher and lower (bracketing) than the standard—zero Asymmetry.

Effect of Context on Judgments

Fig. 2 shows the relation between degree of asymmetry (or, the percentage of stimuli greater than the standard) and the percentage of pitch judgments of 'higher.' Fig. 3 shows this relation for loudness.

The effect of context on the judgment, as displayed in these figures, may be summarized as follows:

1. The percent of 'higher' (or 'louder') judgments increases as the asymmetry becomes more positive,

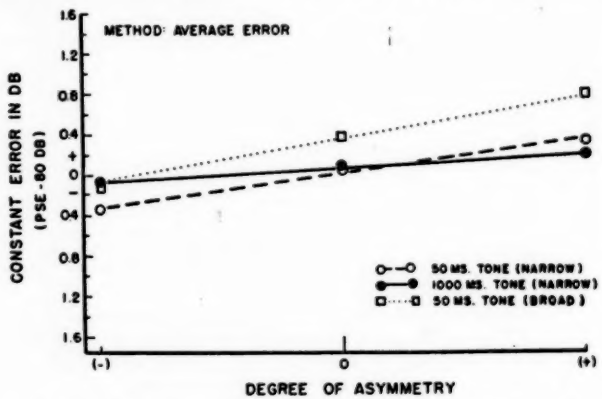


FIG. 7. The CE for loudness as determined by the Method of Average Error as a function of asymmetry. The parameters are the same as those in Fig. 6.

and decreases as the asymmetry becomes more negative.

The points for the 50-msec. tone (open circles) in Fig. 2 show that 36 percent of the pitch judgments were judgments of 'higher' at −3 Asymmetry (or, when 7.1 percent of the stimuli in the context were higher than the standard). At −1 Asymmetry (when 35.7 percent of the stimuli were higher), 45 percent of the judgments were 'higher.'

2. The effect of context increases with increasing asymmetry.

More judgments are influenced by the context under the extreme degrees of asymmetry than under moderate degrees. For instance, in Fig. 2 the ratio of 'higher' judgments to higher stimuli at −3 Asymmetry is 0.20, while this ratio at −1 Asymmetry is 0.79. Since a ratio of 1.00 would indicate no effect of context and one of 0.00 would indicate a maximum effect, the ratio at −3 Asymmetry shows considerably more effect of context at this degree of asymmetry than at −1 Asymmetry.

3. Context affects pitch and loudness judgments equivalently.

Comparing the slopes of equivalent curves (e.g., the 50-msec. curves) in Figs. 2 and 3, it is apparent that their slopes are about equal. This means, of course, that at any degree of asymmetry, about the same percentage of judgments that are called 'higher' in the case of pitch are called 'louder' in the case of loudness.

4. The effect of context is greater on the shorter tone than on the longer tone.

The slope of the 50-msec. curve in Fig. 2 is less than the slope of the 1000-msec. curve. This is also true in Fig. 3. It will be recalled that in these graphs the closer to 0.00 that the slope approaches, the greater is the effect of asymmetry in the context. This is to say that the closer to zero the slopes are, the less the judgments conform to the distribution of the comparison stimuli.

5. The effect of context tends to decrease when the spread of the comparison stimuli is increased beyond a certain point.

Fig. 2 shows that the percent of judgments of 'higher' for the 50-msec. tone at +2 Asymmetry is greater when the stimuli are spread over a relatively broad range (unfilled square) than when spread over a relatively narrow range (unfilled circle). In other words, the distribution of judgments is more in conformity with the distribution of stimuli in the context when the range is broad than when it is narrow. (The limits of the narrow range are defined by the point at which the *O* can make 100 percent accurate judgments. The broad range is twice as wide as the narrow.)

Effect of Context on CE's

Method of Constant Stimuli.—Figs. 4 and 5 show the relation between the CE's for pitch and loudness, respectively, and the arithmetic mean of the comparison stimulus series. (Actually, the arithmetic mean minus the standard.)

Several conclusions about this relationship can be drawn from these figures.

1. The sign of the CE's for pitch and loudness tends to be in the direction in which the context deviates from the standard.

Whenever the context is shifted in the negative direction (when the arithmetic mean of the context is less than the standard), the CE tends to be negative in sign. In Fig. 4, for instance, when the mean of the series is 2.67 cps less than the standard, the

CE for pitch of a 1000-msec. stimulus is −1.3 cps. When the mean of the context is +1.33 cps, the CE is +0.9 cps.

2. The magnitude of the shift in the CE approximates a constant proportion of the physical displacement of the context from the standard.

The ratio between the CE and the difference between the arithmetic mean of the comparison stimulus series minus the standard tends to be constant. This relationship is a more consistent one for loudness (Fig. 5) than for pitch (Fig. 4).

3. The effect of context on the CE is greater for the shorter stimulus than for the longer stimulus.

The ratio between the CE and the displacement of the context from the standard is larger for the shorter (50-msec.) stimulus than for the longer (1000-msec.) one. Comparing the slopes of the 50-msec. and 1000-msec. tones in Fig. 4, it is evident that the slope of the 50-msec. curve is closer to unity than that of the 1000-msec. curve. This means, of course, that the CEs for the shorter tone are closer in magnitude to the difference between the mean of the context and the standard than are the CE's for the longer tone.

4. The constant relation between the size of the CE and the magnitude of the displacement of the context from the standard begins to break down as the displacement goes beyond a certain point.

In Fig. 4, this ratio drops from about 0.41 at +5.33 cps displacement, to about 0.18 at +10.67 cps displacement. In the case of loudness, Fig. 5, however, the ratio remains about the same even at the most extreme displacement at which we measured the effect. We would expect, however, that at some point the ratio would fall off as in the case of pitch.

Method of Average Error.—Figs. 6 and 7 display the effect of asymmetry in the stimulus context on the CE's for pitch and loudness as it occurs in the Method of Average Error. The information provided by these graphs may be summarized as follows:

1. The CE's for pitch and loudness are shifted in the direction of the asymmetry in the context. The amount of shift is about equivalent for positive and negative asymmetries.

The 50-msec. curve for pitch in Fig. 6 shows that the +0.9 cps CE at zero Asymmetry has been shifted to −2.3 cps at (−) Asymmetry and to +4.7 cps at (+) Asymmetry.

2. The effect of context is greater on the shorter tone than on the longer.

The slope for the 50-msec. curves in both Figs. 6 and 7 is steeper than the slope of the 1000-msec. curves indicating that the CE has been increased more for the former than for the latter at (+) and (−) Asymmetry from its value at zero Asymmetry.

3. Increasing the range over which the comparison stimulus can be varied does not increase the effect of asymmetry in the context.

Reference to Figs. 6 and 7 shows that the 50-msec. curves for the narrow range (broken line) and broad range (dotted line) are essentially parallel. This indicated that the change in CE from the condition at zero Asymmetry is about the same for both ranges.

Effect of the Psychophysical Method on CE's

In this section, we are concerned only with the CE's obtained under zero Asymmetry, since we are interested in comparing the methods with respect to CE's as they are obtained

normally. In the usual application of these methods, the context is a balanced one, and, hence, equivalent to our zero Asymmetry.

Pitch.—To compare the two methods with respect to the CE for pitch, see Figs. 4 and 6. When this comparison is made on the basis of the values at zero Asymmetry, the following becomes clear:

1. For the short tone (50-msec.), there is a significant difference in the CE for pitch as determined by the two different psychophysical methods.

As determined by the Method of Constant Stimuli, the CE for pitch of a 50-msec. tone is −2.1 cps. As determined by the Method of Average Error, this value is +0.9 cps.

2. For the long tone (1000-msec.), there is no significant difference in the CE for pitch as determined by the two methods.

Loudness.—To compare the two methods with respect to the CE for loudness, see Figs. 5 and 7. Such a comparison reveals that there is no significant difference in the CE for loudness as determined by the two methods either in the case of the long or short tone.

Effect of Context on DLs

Figs. 8 and 9 show the effect of asymmetry in the stimulus context on the DLs for pitch and loudness as determined by the Method of Constant Stimuli. Figs. 10 and 11 show this effect as determined by the Method of Average Error. The conclusions to be drawn from these graphs follows.

1. The DL's for pitch and loudness are unaffected by asymmetry in the stimulus context.

This generalization holds for both methods and for both long and short tones. In Fig. 8 the tendency for the limen for pitch to increase at negative asymmetries in the case of the 50-msec. stimulus is due mainly to the effect of one *O*'s data on the average. Similarly, this tendency in the limen for loudness (see 1000-msec. curve in Fig. 9) is due to a single *O*.

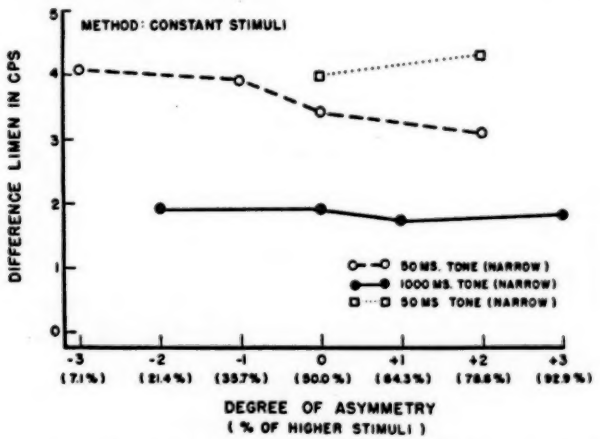


FIG. 8. The DL for pitch as determined by the Method of Constant Stimuli as a function of asymmetry. The parameters are the same as those in Fig. 6.

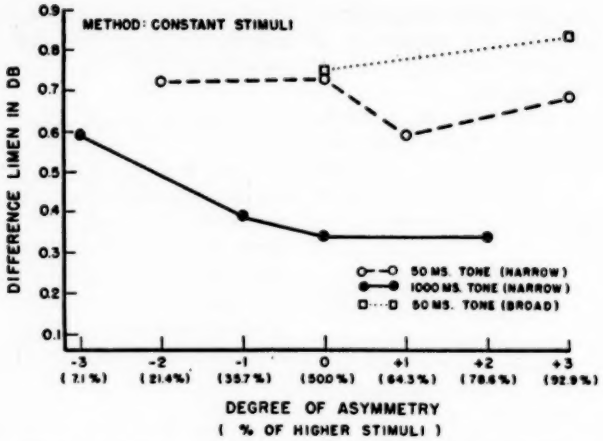


FIG. 9. The DL for loudness as determined by the Method of Constant Stimuli as a function of asymmetry. The parameters are the same as those in Fig. 6.

2. All DL's are increased in size by increasing the spread of the comparison stimuli.

In any of the graphs from Figs. 8 to 11, the values of the points marked by the unfilled squares (the condition of a relatively broad range of comparison stimuli) are greater than those marked by the unfilled circles (the condition of narrow range).

3. Even when the range of the comparison stimuli is increased to a relatively broad one, there is still no demonstrable effect of asymmetry on the DL.

In Fig. 8, for example, the difference in the value of the DL for pitch at +2 Asymmetry and at zero Asymmetry, under the condition of an increased range of stimuli (unfilled

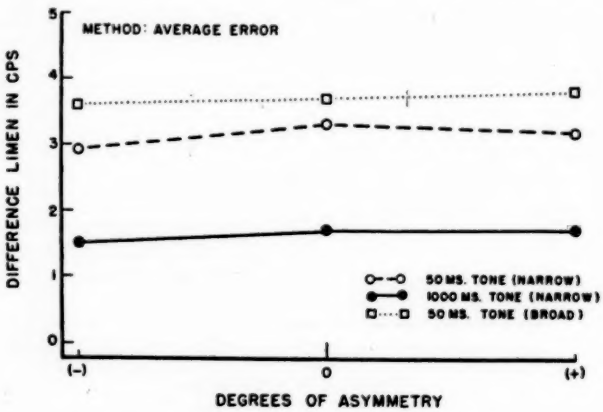


FIG. 10. The DL for pitch as determined by the Method of Average Error as a function of asymmetry. The parameters are the same as those in Fig. 6.

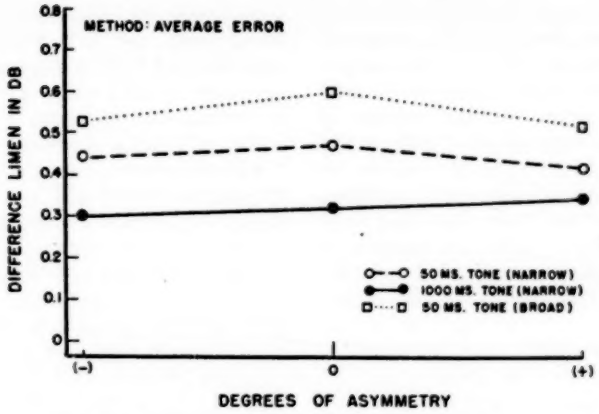


FIG. 11. The DL for loudness as determined by the Method of Average Error as a function of asymmetry. The parameters are the same as those in Fig. 6.

squares), is not a significant one. Comparing the values at zero Asymmetry and +3 Asymmetry in Fig. 9 under the condition of an increased range of stimuli (unfilled squares) reveals no significant difference in the DL for loudness. This same conclusion applies for Figs. 10 and 11.

DISCUSSION

Implications for Methodology

Determination of CE's: Method of Constant Stimuli.—This experiment amply demonstrates that the psychophysical judgment and, consequently, the CE is seriously affected by the asymmetries in the stimulus context which we introduced. The consistency of this effect from *O* to *O* and the degree to which the CE is affected leads to the conclusion that the Method of Constant Stimuli should not be used for the determination of CEs.

Two of our *O*s were experimentally naive. The other was not only experienced in psychological experiments but had developed the reputation of being extremely skeptical of everything that went on in such experiments. In addition, he had the keenest discrimination of the three *O*s. (His DL for pitch was on the order of 1.1 cycles for a 1000 cps, 1000 msec. tone.) Even this *O*'s judgments were influenced by the asymmetries and to a degree that his results are essentially the same as those of the other two *O*s. The generalization that one is led to in view of these considerations is that the effect of context is probably a universal one.

The degree to which the CE can be influenced is quite large, and is directly related to the actual size of the CE. To make this clear, let us take an example. Suppose our problem is to determine the effect of duration on loudness, using the Method of Constant Stimuli. Suppose, too, that at 50-msec. there is an actual CE of −1.6 db. We do not know this is advance, of course, since to determine it is the purpose of the experiment. If we follow the usual procedure, we will select our comparison stimuli so that we have a balanced (physically) series about the standard. The unknown (to us) CE, however,

makes our series unbalanced insofar as the experience of the *O* is concerned. In other words, we have introduced some degree of asymmetry in the context. According to the results of this study, our experimentally determined CE is going to be in error by about 50 percent of the extent to which the context has been displaced from true symmetry (i.e., 50 percent of 1.6 db). This error will be in the direction of reducing the size of the CE.

This particular error amounts to about one JND (Fig. 9 shows that the DL for loudness at 50-msec. duration is about 0.73 db). If the actual CE was −5 db the experimentally determined CE would be in error by fully three JNDs.

The extent to which context influences the CE also seems to depend upon the difficulty of the task of discrimination. For instance, the effect of context is greater for the 50-msec. tone as for the 1000-msec. tone. Apparently, there is as much pitch or loudness in a 50-msec. tone as there ever will be, but the total experience of the sound is sufficiently different from the experience of the standard to make the discrimination more difficult than it is when the two are of equivalent duration and, hence, more alike in every respect. We would predict that by making the task of discrimination even more difficult (shortening the duration still further) we would increase the effect of context.

Determination of CE's: Method of Average Error.—We cannot state categorically that this method is free from context effects. There is some evidence from this experiment, however, which would indicate that the judgment may be less affected by context in this method than in the Method of Constant Stimuli. Figs. 6 and 7 show that the degree to which the CE is displaced by reason of either (−) Asymmetry or (+) Asymmetry is approximately equivalent to the DL (see Figs. 10 and 11). This appears to be the case for both the narrow and broad ranges of stimuli.

This evidence suggests that the displacement of the CE is due to an adjustment tendency on the part of the *O*. For (−) and (+) Asymmetries, the *O* was instructed to move the variable stimulus in one direction only until he had made the match. It is possible that the *O* tended to move the variable stimulus up to the edge of the interval of uncertainty (the limit of which on one side of the standard is defined by the DL) and to stop there. This interpretation is strengthened by the fact that broadening the range over which the *O* had to vary the comparison stimulus did not effect the degree to which the CE was displaced any more than did the condition of a relatively narrow range. The context was altered but the displacement of the CE remained the same.

This explanation does not, of course, prove that there are no context effects in the Method of Average Error. The superiority of this method in determining CE's does not stand or fall on this point. Rather, we believe that the determination of the context by the *O* results in a context which is likely to be balanced psychologically. In this method the *O* is usually instructed to move the variable stimulus alternately above and below the standard, gradually closing in on the standard. Since these adjustments can be made by the *O* only on the basis of the quality in question (pitch or loudness), any CE's are taken into account in the *O*'s adjustment of the variable. Consequently, the context tends to be a balanced one insofar as the experience of the *O* is concerned.

Determination of DL's.—The results of this experiment reveal that the DL is unaffected by context and that the Method of Constant Stimuli and the Method of Average Error give essentially the same DL's.

In view of these findings, it would appear that either method is suitable for the determination of the DL.

Nature of the Context Effects

This experiment was not set up to answer the question of the nature of the context effects on judgments and, hence, CE's. However, we would like to give a brief summary of some of the kinds of explanations of context effects which have been given in the past and which might apply to our data.[3] We would also like to add one of our own.

Koester and Schoenfeld (6) recently had the occasion to review a number of different hypotheses concerning the nature of the interaction effects occurring between members of a stimulus series.

1. Hollingsworth some time ago attributed to the *O* a tendency to make his judgments with respect to the median value of the stimulus series with which the *O* was concerned.

2. Woodrow thought of the interaction effects in terms of a tendency for the standard stimulus to be assimilated toward the weighted average of the stimulus series, with obvious effects on the judgments made with respect to this standard.

In an effort to explain the results of their own experiment, Koester and Schoenfeld suggested a few other possible tendencies on the part of the *O:*

3. The tendency for the *O* to make absolute judgments—calling stimuli 'high' or 'low' (or 'loud' or 'soft') depending upon the position of the stimulus in the scale determined by the particular series of stimuli in use during the run.

4. The tendency for the *O* to make judgments with respect to the preceding comparison rather than the present standard.

5. The tendency for the present standard to be contrasted with the preceding levels of stimulation, with consequent effects on the judgment of the comparison stimulus that follows.

6. The possibility of several of these tendencies working in combination.

No one of these explanations is adequate to explain our results completely. (Nor was any one adequate to explain all of Koester and Schoenfeld's data, as they pointed out.)

There is another possibility, however, which ought to be mentioned.

7. There is a tendency for the *O* to split his judgments 50–50 whenever presented with a difficult two-choice situation (as in our experiment).

Our results are at least partly interpretable in terms of such a tendency. As asymmetry increases, there is an increasing paucity of one of the categories of response on the *O*'s recording sheet. He tends, therefore, as this paucity increases to give more and more of the other category of response when the discriminations are difficult. This tendency would account for the growing disparity between the actual distribution of the

[3] Helson explains these effects in terms of an adaptation level which comes to constitute the reference in terms of which the judgments of the *O* are made, and which the PSE approximates. The adaptation level is determined by all the stimuli in the series to which the *O* is subjected, with the greatest influence being contributed by the comparison stimuli. Helson provides formulae for the computation of this adaptation level, and, hence, the prediction of the PSE. See Helson, H. Adaptation level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, 60, 1–29.

stimuli and the distribution of judgments with increasing asymmetry.

Such an interpretation is supported by the fact that when the range over which the comparison stimuli are spread is increased and the asymmetry kept constant, the distribution of judgments is more like the actual distribution of stimuli than when the range is narrow. In other words, the task of discrimination is easier with the comparison stimuli spread over a wide range and this tendency to split judgments 50–50 is less pronounced. Similarly, one might argue that the task of discrimination is more difficult for the 50-msec. tone than for the 1000-msec. tone and that this accounts for the different distributions of judgments obtained under the same distribution of stimuli.

This explanation, of course, does not account for the absolute values of the CE which were obtained. These values, it will be remembered, seem to be more closely related to the displacement of the physical context from the physical standard (arithmetic mean of the series minus the standard) than to the degrees of asymmetry (distribution of stimuli).

Summary and Conclusions

This experiment was designed to test the prediction that the *E* can predetermine the CE's for pitch and loudness by the selection of the comparison stimuli in the Method of Constant Stimuli. Toward this end CE's were determined under varying conditions of asymmetry in the comparison stimulus series. The CEs for both pitch and loudness were found to vary in a regular manner with these asymmetries in the context. The degree to which the size of the CE could be influenced was found to be quite large. Similar effects were seen to occur in the Method of Average Error. These results are discussed and lead to the following conclusions:

1. Because of the arbitrary selection of the context by the *E* in the ordinary use of the Method of Constant Stimuli, this method should not be used in the determination of CE's.

2. Because it is the *O* who determines the context within which his judgments are made in the ordinary use of the Method of Average Error, this method is more likely to give true CE's than is the Method of Constant Stimuli.

3. These two methods give essentially the same results insofar as DL's are concerned and, therefore, either would be suitable for the determination of DL's.

(Manuscript received for immediate publication July 25, 1949)

References

1. Doughty, J. M., & Garner, W. R. Pitch characteristics of short tones. II. Pitch as a function of tonal duration. *J. exp. Psychol.*, 1948, 38, 478–494.
2. Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, 3, 126–150.
3. Gault, R. H., & Goodfellow, L. D. Sources of error in psycho-physical meas, urements. *J. gen. Psychol.*, 1940, 23-197–200.
4. Koester, T. Time-error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level. *Arch. Psychol.*, 1945, 297.
5. Koester, T., & Schoenfeld, W. N. Modified methods of constant and single stimuli in pitch discrimination. *Amer. J. Psychol.*, 1945, 58, 540–544.
6. Koester, T., & Schoenfeld, W. N. The effect of context upon judgments of pitch differences. *J. exp. Psychol.*, 1946, 36, 417–430.
7. Needham, J. G. Contrast effects in judg-

ments of auditory intensities. *J. exp. Psychol.*, 1935, **18**, 214–226.

8. POSTMAN, L. Time-error as a function of the method of experimentation. *Amer. J. Psychol.*, 1947, **60**, 101–108.
9. ROSE, G. L., & HARRIS, J. D. Suggestions for constructing tests of pitch discrimination. Progress Report No. 2. Bureau of Medicine and Surgery, Research Project No. NM 003 026, 10 February 1948.
10. STEVENS, S. S. The volume and intensity of tones. *Amer. J. Psychol.*, 1934, **46**, 397–408.
11. WOODROW, H. Weight-discrimination with a varying standard. *Amer. J. Psychol.*, 1933, **45**, 391–416.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	1	–	3	4	5	6	$1.25	$6.25
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	By subscription, $7.00						$1.25	——
									——
		List price, Volumes 1 through 38							$247.50
		30% discount							74.25
									——
		Net price, Volumes 1 through 38							$173.25

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year. Eighteen numbers affecting eight volumes are out of print in the entire series.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W. Washington 5, D. C.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W.
Washington 5, D. C.

The American Psychological Association, founded in 1892 and incorporated in 1925, is the major psychological organization in the United States. With approximately 6,700 members and 2,000 affiliates, it includes most of the qualified psychologists in the country. The purpose of the APA is to advance psychology as a science, as a profession, and as a means of promoting human welfare. It attempts to further these objectives by holding annual meetings, publishing psychological journals, conducting a personnel placement service for psychologists, and working toward improved standards for psychological training and service.

In order to give recognition to the specialized interests of different psychologists, the APA includes seventeen Divisions. Any person, after becoming a member of the APA, may apply for membership in as many Divisions as he wishes. The Divisions are:

Division of General Psychology
Division on the Teaching of Psychology
Division of Experimental Psychology
Division on Evaluation and Measurement
Division on Childhood and Adolescence
Division of Personality and Social Psychology
The Society for the Psychological Study of Social Issues—a Division of the American Psychological Association
Division on Esthetics
Division of Clinical and Abnormal Psychology
Division of Consulting Psychology
Division of Industrial and Business Psychology
Division of Educational Psychology
Division of School Psychologists
Division of Counseling and Guidance Psychologists
Division of Psychologists in Public Service
Division of Military Psychology
Division on Maturity and Old Age

Each Division has its own officers. Each meets annually at the time and place of the APA meeting. Each has its own membership requirements, which in some cases are higher or more specialized than the requirements for election to the APA.

The annual meetings of the APA are held in September. An attempt is made to choose meeting sites in such a way that attendance will be convenient for members and their guests in different sections of the country at different times. Thus, the 1949 meeting was held on September 6 to 10 in Denver, and the 1950 meeting on September 4 to 11 at Pennsylvania State College. These meetings provide the members with an opportunity to present or hear reports of psychological research and lectures on topics of psychological interest, and to participate in formal and informal discussion groups. The research reports, lectures, and discussion groups cover varied topics of current interest within the field of psychology.

The chief governing body of the APA is the Council of Representatives. It includes representatives from each of the Divisions. A Board of Directors, composed of the four officers of the APA (President, President-elect, Recording Secretary, and Treasurer) and six Council members elected by the Council, is the administrative agent of the Council and exercises general supervision over the affairs of the Association.

The APA maintains an office in Washington which acts as a coordinating center for all APA activities. Journal subscriptions, membership applications, personnel placement requests, and general APA business procedures are handled in this office.